

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 11:14:06 ; Search time 15776 Seconds  
(without alignments)  
11475.903 Million cell updates/sec

Title: US-10-049-137-1  
Perfect score: 4177  
Sequence: 1 atgagtagtttggtggagag.....caactggtaaatcaagattc 4177

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of Hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4177	100.0	4334	8	AF185578	AF185578 Arabidops
2	4156	99.5	4180	8	AF185577	AF185577 Arabidops
3	850.8	20.4	3204	6	AX507831	AX507831 Sequence
4	471.6	11.3	6398	5	BC046866	BC046866 Xenopus 1
5	436.4	10.4	4343	10	BC060721	BC060721 Mus muscu
6	427.4	10.2	6438	10	BC058578	BC058578 Mus muscu
7	423	10.1	6434	9	AK128741	AK128741 Homo sapi
8	421.4	10.1	6327	6	A56817	A56817 Sequence 1
9	421.4	10.1	6328	6	AR224019	AR224019 Sequence
10	421.4	10.1	6328	6	AR269336	AR269336 Sequence
11	421.4	10.1	6328	6	AR367436	AR367436 Sequence
12	421.4	10.1	6417	6	AX332238	AX332238 Sequence
13	421.4	10.1	6417	9	HSMI21218	X86691 H. sapiens m
14	421.4	10.1	6475	6	AR338834	AR338834 Sequence
15	420.8	10.1	6331	6	AX924075	AX924075 Sequence
16	420.8	10.1	6331	6	AX924076	AX924076 Sequence
17	420.8	10.1	6331	9	AF006515	AF006515 Homo sapi
18	420.6	10.1	5689	3	AF308445	AF308445 Caenorhab
19	420.6	10.1	6771	6	AX924049	AX924049 Sequence
20	420.6	10.1	6771	9	HSU91543	U91543 Homo sapien
21	400.2	9.6	6584	9	BC038596	BC038596 Homo sapi
22	387	9.3	3716	6	AX746951	AX746951 Sequence
23	387	9.3	3716	9	AK091386	AK091386 Homo sapi
24	387	9.3	4563	9	HSM801006	AL117491 Homo sapi
25	387	9.3	9646	9	AF425231	AF425231 Homo sapi
26	385.4	9.2	4386	9	HSM802121	AL137693 Homo sapi
27	385.4	9.2	4432	9	HSM800877	AL110281 Homo sapi
28	374	9.0	18029	3	CEH06001	Z92970 Caenorhabdi
29	363.8	8.7	5603	3	AF308444	AF308444 Caenorhab
30	356.6	8.5	4966	9	AK127046	AK127046 Homo sapi
31	347.4	8.3	42467	3	U55373	U55373 Caenorhabdi
32	324.8	7.8	156251	2	AC107325	AC107325 Drosophil
33	324.8	7.8	183948	3	AC009376	AC009376 Drosophil
34	324.8	7.8	194308	3	AC009377	AC009377 Drosophil
35	324.8	7.8	257692	3	AE003517	AE003517 Drosophil
36	322.8	7.7	157766	2	AC146651	AC146651 Medicago
37	321.6	7.7	4912	3	AF007780	AF007780 Drosophil
38	321.2	7.7	89380	8	AP004259	AP004259 Oryza sat
39	318.2	7.6	81729	8	AB016874	AB016874 Arabidops
40	318.2	7.6	82359	8	ATAC002342	AC002342 Arabidops
41	317.2	7.6	92783	3	AC084459	AC084459 Caenorhab
42	313.2	7.5	6475	3	AF119716	AF119716 Drosophil
43	312.8	7.5	3079	3	AY071503	AY071503 Drosophil
44	311.6	7.5	150884	2	AC009382	AC009382 Drosophil
45	311.6	7.5	167816	3	AC009381	AC009381 Drosophil

ALIGNMENTS

RESULT 1  
AF185578  
LOCUS AF185578 4334 bp mRNA linear PLN 10-NOV-1999  
DEFINITION Arabidopsis thaliana GYMNO5/PICKLE mRNA, complete cds.  
ACCESSION AF185578  
VERSION AF185578.1 GI:6318929  
KEYWORDS  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 4334)  
REFERENCE Eshed,Y., Baum,S.F. and Bowman,J.L.  
AUTHORS Distinct mechanisms promote polarity establishment in carpels of  
TITLE

JOURNAL	Arabidopsis thaliana
REFERENCE	Cell (1999) In press
AUTHORS	2 (bases 1 to 4334)
TITLE	Eshed, Y.
JOURNAL	Direct Submission
FEATURES	Submitted (12-SEP-1999) Plant Biology, UC Davis, One Shields Ave., Davis, CA 95616, USA
source	Location/Qualifiers
	1..4334
	/organism="Arabidopsis thaliana"
	/mol_type="mRNA"
	/cultivar="Colombia"
	/db_xref="taxon:3702"
	/chromosome="2"
	5..4159
CDS	/note="Mi-2 like; CHD chromatin remodeling enzyme homolog"
	/codon_start=1
	/product="GYMNOS/PICKLE"
	/protein_id="AAF07084.1"
	/db_xref="GI:6318930"
	/translation="MSSLVERLRIIRSDRKPVYNLDDSDDDFVPKKDRTFEQVEAIVR
	TDAKENACQAGESTNLVSCNTCTYAFHAKCLVPPLKDSAVENWRCPECVSPLEIDK
	ILDCMERPTKSESQSSDAEPKPIFVKQYLVKWKGLSYLHGSVWVPEKEFKAYKSNHR
	LKTRVNNFHRQMESFNNSDEDDFVAIRPEWTTVDRIACREDEGELEYLVKYKELSYDE
	CWYESDISTSFQNEIQRFKDVNSRTRRSKDVHKNRPRDFOQDHTPEFLKGLLHPY
	QLEGSLNFRFSWKQTHVILADENGLGKTIQSIALLASLFEENLPHLVIAPLSLTLRN
	WEREFATWAPQMNVVMYFGTAQARAVIREHEFYLSKDQKTKKKSKSQISSSESKQRI
	KFDVLLTSYEMINLDSAVLKPIKWECMIVDEGHLRKNKDSKLFSSLTQYSSNHRILLT
	GTPLQNNLDELFLMHFLDAGKFGSLSEFQEFKDINQEEQISRLHKMLAPHLRRVK
	KDVMKMPKPKELILRLVDLSLQKEYYKAIFTRNYQVLTFRKGAQISLNNIMMELRKV
	CCHPYMLEGVEPVIHDANEAFKQLLESCKLQLLDKMVKLKEQHRVLIYTFQFHLML
	DLLEDYCTHKWQYERIDGKVGGAERQIRIDRFNAKNSNKFCLLSSTRAGGLGINLAT
	ADTVIIYDSDNPHADLQAMARHRLQOTNKVMIYRLINRGTEIERMMQLTKKMWLE
	HLVVGKLTQNIQEEELDDIIRYGSKELFASDEDEAGSKGKIHYDDAAIDKLILDRDLV
	EABEVSVDDEEENGFLKAFKANFEYIDENEAAALEAQRVAAESKSSAGNSDRASYME
	ELLKDKFELHQAELNALGKRKRSRQLVSI EEDDLAGLEDVSDGDES YEAESTDGE
	AAGQGVQTRRPYRRKGRDNLEPTPLMEGEGRSFRVLGNQSORAIFVQTLRMRYGAGN
	FDWKEFVRLKQKTFEEINEYGILFLKHIAEEIDENSPTFSDGVPKEGLRIEDVLVRI
	ALLILVQEKVKFVEDHPGKVPFSPRIILERPPGLRSGKIWKEBHDKIMIRAVLKHGYGR
	WQAIVDKELGIQELICKELNFPHISLSAAEQAGLQONGSGSGSNPQAOTNQNPGSVI
	TGNNASADGAQVNSMFYRDMORRLVEFVKRVLLEKAMNYEYAEYYGLGSSSI
	PTEEPEAPFKIADTVGSFIEVDDEMLDGLPKTDPIITSEEINGAAVDNNQARVEIAQH
	YNQMKLILDENARESVOQYVNNQPPSTKVNESFRALKSINGNINTILSITSDQKSHE
	DDTKPDLNNVEMKDTAEETKPLRGVVDLNVNVEGEENIAEASGSDVDVKMEEAKEEEKP
	KNMVVD"
ORIGIN	
Query Match	100.0%; Score 4177; DB 8; Length 4334;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 4177; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 ATGAGTAGTTTGGTGAGAGGCTTCGCATACGATCTGATAGGAAACCAGTTTATAACCTA 60
Db	5 ATGAGTAGTTTGGTGAGAGGCTTCGCATACGATCTGATAGGAAACCAGTTTATAACCTA 64
QY	61 GATGATTCTGATGATGACGACTTCGTTCCCTAAAAAAGATCGAACCTTTGAGCAAGTCGAG 120
Db	65 GATGATTCTGATGATGACGACTTCGTTCCCTAAAAAAGATCGAACCTTTGAGCAAGTCGAG 124
QY	121 GCTATTGTCAGAACTGATGCGAAAGAAAATGCAATGTCAGGCTTGTGGGAAAAGTACTAAT 180
Db	125 GCTATTGTCAGAACTGATGCGAAAGAAAATGCAATGTCAGGCTTGTGGGAAAAGTACTAAT 184
QY	181 CTTGTAAGCTGCAATACATGCACCTTATGCGTTTCATGCTAAATGCTTAGTCCACCTCTT 240
Db	185 CTTGTAAGCTGCAATACATGCACCTTATGCGTTTCATGCTAAATGCTTAGTCCACCTCTT 244
QY	241 AAAGATGCTTCGGTGGAAAAATTGGAGATGCCCTGAAATGTTAGTCTCTTAACGAGATA 300
Db	245 AAAGATGCTTCGGTGGAAAAATTGGAGATGCCCTGAAATGTTAGTCTCTTAACGAGATA 304
QY	301 GATAAGATATTGGATTGTGAAATGCGTCCCTACAAAATCTAGTGAACAAGGTTTCCTCCGAT 360

Db	305 GATAAGATATTGGATTGTGAAATGCGTCCTTACAAAATCTAGTGAACAAGGTTCTCTCCGAT 364
QY	361 GCGGAACCGAAGCCAAATTTTGTGAAACAGTATCTCGTGAAGTGAAGGGATTATCATAC 420
Db	365 GCGGAACCGAAGCCAAATTTTGTGAAACAGTATCTCGTGAAGTGAAGGGATTATCATAC 424
QY	421 CTTCACTGCTCTTGGGTGCTGAGAGGAGTTCAGAAAGGCTTATAAGTCAAATCATCGT 480
Db	425 CTTCACTGCTCTTGGGTGCTGAGAGGAGTTCAGAAAGGCTTATAAGTCAAATCATCGT 484
QY	481 TTAAAAACAGAGTGAAACAATTTTCACCGTCAAATGGAGTCAATCAATAACAGCGAAGAT 540
Db	485 TTAAAAACAGAGTGAAACAATTTTCACCGTCAAATGGAGTCAATCAATAACAGCGAAGAT 544
QY	541 GATTTTGTGCCATACGTCCTGAGTGGACCACTGTTGATCGGATCTTTGCCTGCAGAGAG 600
Db	545 GATTTTGTGCCATACGTCCTGAGTGGACCACTGTTGATCGGATCTTTGCCTGCAGAGAG 604
QY	601 GAAGATGGGAGCTGGAATATCTTGTCAAATATAAAGAGTATCCTATGATGAATGTTAT 660
Db	605 GAAGATGGGAGCTGGAATATCTTGTCAAATATAAAGAGTATCCTATGATGAATGTTAT 664
QY	661 TGGGAGTCAGAAATCAGACATCTCAACCTTCCAGAAATGAAATTCAAAGGTTCAAGGATGTA 720
Db	665 TGGGAGTCAGAAATCAGACATCTCAACCTTCCAGAAATGAAATTCAAAGGTTCAAGGATGTA 724
QY	721 AATTCTAGAACTCGCAGAAAGTAAAGATGTTGACCATAAAAGAAATCCAGAGACTTTCAA 780
Db	725 AATTCTAGAACTCGCAGAAAGTAAAGATGTTGACCATAAAAGAAATCCAGAGACTTTCAA 784
QY	781 CAGTTTGATCATACTCTGAAATTCCTCAAAGGCTTGTACATCCATACCAGCTTGAGGGA 840
Db	785 CAGTTTGATCATACTCTGAAATTCCTCAAAGGCTTGTACATCCATACCAGCTTGAGGGA 844
QY	841 CTTAATTTTGTGGGTTCTCGTGTCAAATAACAGACGATGTAATCTTGTCTGATGAAATG 900
Db	845 CTTAATTTTGTGGGTTCTCGTGTCAAATAACAGACGATGTAATCTTGTCTGATGAAATG 904
QY	901 GGACTAGGCAAGACAATTCAAAGCATGCCCCTTTTAGCTTCACTTTTGGAGGAGAACCTC 960
Db	905 GGACTAGGCAAGACAATTCAAAGCATGCCCCTTTTAGCTTCACTTTTGGAGGAGAACCTC 964
QY	961 ATTCCGCAATTGGTAATTGCTCCTCTATCGACTCTCGCTAAGTGGGAGAGAGAGTTTGCC 1020
Db	965 ATTCCGCAATTGGTAATTGCTCCTCTATCGACTCTCGCTAAGTGGGAGAGAGAGTTTGCC 1024
QY	1021 ACATGGGCCCCACAGATGAACGTTGTTATGTATTTTGGCACTGCGCAAGCTCGAGCAGTT 1080
Db	1025 ACATGGGCCCCACAGATGAACGTTGTTATGTATTTTGGCACTGCGCAAGCTCGAGCAGTT 1084
QY	1081 ATCAGAGAACATGAGTTTTTACTTATCGAAAGATCAAAAAAAGATCAAGAAAAAGAAATCT 1140
Db	1085 ATCAGAGAACATGAGTTTTTACTTATCGAAAGATCAAAAAAAGATCAAGAAAAAGAAATCT 1144
QY	1141 GGACAAATAAGTAGCGAAAGCAAGCAAAAAAGAAATCAAGTTTGATGTCCTCCTCACATCG 1200
Db	1145 GGACAAATAAGTAGCGAAAGCAAGCAAAAAAGAAATCAAGTTTGATGTCCTCCTCACATCG 1204
QY	1201 TATGAGATGATCAACCTAGATTGAGCAGTTCTAAAACCAATTAAGTGGGAGTGCATGATT 1260
Db	1205 TATGAGATGATCAACCTAGATTGAGCAGTTCTAAAACCAATTAAGTGGGAGTGCATGATT 1264
QY	1261 GTTGATGAAGGTCATCGACTGAAAAATAAGGATTCAAAGCTGTTCTCTTCAATTGACACAG 1320
Db	1265 GTTGATGAAGGTCATCGACTGAAAAATAAGGATTCAAAGCTGTTCTCTTCAATTGACACAG 1324
QY	1321 TATTCAAGTAACCAACCGTATTCTTCTGACAGGAACACCACCTTCAGAACAACTTGGATGAA 1380
Db	1325 TATTCAAGTAACCAACCGTATTCTTCTGACAGGAACACCACCTTCAGAACAACTTGGATGAA 1384
QY	1381 CTTTTCATGCTCATGCATTTTCTTGATCGGGGGAAGTTTGGAGGAGTTTCCAG 1440
Db	1385 CTTTTCATGCTCATGCATTTTCTTGATCGGGGGAAGTTTGGAGGAGTTTCCAG 1444



QY 1441 GAGGAGTTCAAAGATATTAAATCAAGAGGAGCAGATCTCAAGGTTGACAAAAATGTTGGCT 1500  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
1445 GAGGAGTTCAAAGATATTAAATCAAGAGGAGCAGATCTCAAGGTTGACAAAAATGTTGGCT 1504  
QY 1501 CCACATTTGCTCAGAAGGGTAAAAAAGACGTAATGAAAAGACATGCCCCCAAAAAGGAG 1560  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
1505 CCACATTTGCTCAGAAGGGTAAAAAAGACGTAATGAAAAGACATGCCCCCAAAAAGGAG 1564  
QY 1561 CTCATTTGCGTGTGATCTGAGCAGTCTGCAGAAAAGAAATATTACAAAAGCTATTTTACC 1620  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
1565 CTCATTTGCGTGTGATCTGAGCAGTCTGCAGAAAAGAAATATTACAAAAGCTATTTTACC 1624  
QY 1621 CGTAATTATCAAGTATTGACAAAAAAGGGAGGTGCTCAAAATTTCCCTTAATAACATTATG 1680  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
1625 CGTAATTATCAAGTATTGACAAAAAAGGGAGGTGCTCAAAATTTCCCTTAATAACATTATG 1684  
QY 1681 ATGGAATTACGAAAAGTATGCTGCCATCCTTATATGCTAGAGGGTGTGAGCCAGTTATT 1740  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
1685 ATGGAATTACGAAAAGTATGCTGCCATCCTTATATGCTAGAGGGTGTGAGCCAGTTATT 1744  
QY 1741 CACGACGCAAAATGAAGCTTTCAAAACAATTTTGGAGTCTTGTGGAAGCTGCAACTTCTA 1800  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
1745 CACGACGCAAAATGAAGCTTTCAAAACAATTTTGGAGTCTTGTGGAAGCTGCAACTTCTA 1804  
QY 1801 GATAAAATGATGCTCAAACTGAAAGAGCAAGGACACAGAGTCCTAATATACACACAGTTT 1860  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
1805 GATAAAATGATGCTCAAACTGAAAGAGCAAGGACACAGAGTCCTAATATACACACAGTTT 1864  
QY 1861 CAGCATATGCTGGACTTACTTTGAAGACTACTGTACCCATAAGAAATGGCAGTACGAGCGA 1920  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
1865 CAGCATATGCTGGACTTACTTTGAAGACTACTGTACCCATAAGAAATGGCAGTACGAGCGA 1924  
QY 1921 ATTGATGGAAGGTTGGCGGAGCTGAGCGGCAAAATACGCATAGATCGGTTCAATGCCAAA 1980  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
1925 ATTGATGGAAGGTTGGCGGAGCTGAGCGGCAAAATACGCATAGATCGGTTCAATGCCAAA 1984  
QY 1981 AATTCTAACAGTTTGTGTTTTGCTCTCCACAAGAGCTGGTGGCTTAGGAATAAAATCTT 2040  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
1985 AATTCTAACAGTTTGTGTTTTGCTCTCCACAAGAGCTGGTGGCTTAGGAATAAAATCTT 2044  
QY 2041 GCAACGGCTGATACAGTAATCAATTTATGACAGTGACTGGAATCCTCATGCTGATCTTCAA 2100  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2045 GCAACGGCTGATACAGTAATCAATTTATGACAGTGACTGGAATCCTCATGCTGATCTTCAA 2104  
QY 2101 GCAATGGCTAGAGCTCATCGACTTGGCCAAAACAAATAAGGTGATGATTTATAGGCTCATA 2160  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2105 GCAATGGCTAGAGCTCATCGACTTGGCCAAAACAAATAAGGTGATGATTTATAGGCTCATA 2164  
QY 2161 AACCGAGGCACCATTTGAAGAAAGGATGATCAATTGACTAAAAAGAAAAATGTTCTAGAG 2220  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2165 AACCGAGGCACCATTTGAAGAAAGGATGATGCAATTGACTAAAAAGAAAAATGTTCTAGAG 2224  
QY 2221 CATCTTGTGTTGGGAAACTCAAAACACAAAACATTAATCAGGAAGAGTTAGATGACATC 2280  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2225 CATCTTGTGTTGGGAAACTCAAAACACAAAACATTAATCAGGAAGAGTTAGATGACATC 2284  
QY 2281 ATCAGGTATGGATCAAAAGGAGCTTTTGTGCTAGTGAAGATGATGAAGCAGGAAAGTCTGGA 2340  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2285 ATCAGGTATGGATCAAAAGGAGCTTTTGTGCTAGTGAAGATGATGAAGCAGGAAAGTCTGGA 2344  
QY 2341 AAAATTCAATTATGATGATGCGGCTATAGACAAAATGCTTGTATCGTATCGTGGAGGCA 2400  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2345 AAAATTCAATTATGATGATGCGGCTATAGACAAAATGCTTGTATCGTATCGTGGAGGCA 2404  
QY 2401 GAGGAAGTCTCAGTGGATGATGAAGAGGAGAAATGGATTCTTAAAGGCTTCAAGGTGGCT 2460  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2405 GAGGAAGTCTCAGTGGATGATGAAGAGGAGAAATGGATTCTTAAAGGCTTCAAGGTGGCT 2464  
QY 2461 AATTTTGAATATATAGATGAAAATGAGGCAGCAGCATTAGAGGCAACAGAGAGTCGCTGCT 2520  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2465 AATTTTGAATATATAGATGAAAATGAGGCAGCAGCATTAGAGGCAACAGAGAGTCGCTGCT 2524

QY 2521 GAAAGCAAATCTTCAGCAGGCAATTTCTGATAGAGCAAGTTATTGGAAGAGTTGTTAAAA 2580  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2525 GAAAGCAAATCTTCAGCAGGCAATTTCTGATAGAGCAAGTTATTGGAAGAGTTGTTAAAA 2584  
QY 2581 GATAAATTTGAGCTGCACCAGGCTCAGGAGCTTAATGCTCTTGGAAAAAGGAAGAAGT 2640  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2585 GATAAATTTGAGCTGCACCAGGCTCAGGAGCTTAATGCTCTTGGAAAAAGGAAGAAGT 2644  
QY 2641 CGCAAGCAGTTGGTATCCATTGAAAGAGATGATCTTGTGCTGGTTTGGAAAGATGTGAGCTCT 2700  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2645 CGCAAGCAGTTGGTATCCATTGAAAGAGATGATCTTGTGCTGGTTTGGAAAGATGTGAGCTCT 2704  
QY 2701 GATGGAGTGAAGTATTATGAAGCTGAGTCAACAGATGGTGAAGCAGGACAAAGGAGTT 2760  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2705 GATGGAGTGAAGTATTATGAAGCTGAGTCAACAGATGGTGAAGCAGGACAAAGGAGTT 2764  
QY 2761 CAGACGGTGCACGGCCGTACAGAAAGAAAGGTCGCGATAATTTGGAAACCAACTCCGTTG 2820  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2765 CAGACGGTGCACGGCCGTACAGAAAGAAAGGTCGCGATAATTTGGAAACCAACTCCGTTG 2824  
QY 2821 ATGGAAGGTGAGGGGAGATCTTTTTCAGAGTACTGGGTTTCAACCAGAGTCAAAAGGGCCATT 2880  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2825 ATGGAAGGTGAGGGGAGATCTTTTTCAGAGTACTGGGTTTCAACCAGAGTCAAAAGGGCCATT 2884  
QY 2881 TTTGTACAGACTTTTGTATGAGGTATGGAGCTGGCAATTTTGTATTGGAAGGAGTTTGTTCCT 2940  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2885 TTTGTACAGACTTTTGTATGAGGTATGGAGCTGGCAATTTTGTATTGGAAGGAGTTTGTTCCT 2944  
QY 2941 CGTTAAAGCAGAAGACCTTTTGAAGAAATAAAATGAATATGGAATACTCTTCTTGAAGCAC 3000  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2945 CGTTAAAGCAGAAGACCTTTTGAAGAAATAAAATGAATATGGAATACTCTTCTTGAAGCAC 3004  
QY 3001 ATTGCTGAAGAAATAGACGAGAAATTTCTCCAACTTTTTCAGATGGTGTGCCCAAGGAAGGA 3060  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
3005 ATTGCTGAAGAAATAGACGAGAAATTTCTCCAACTTTTTCAGATGGTGTGCCCAAGGAAGGA 3064  
QY 3061 CTTAGAAATAGAAGATGTTCTAGTCAGAAATGCTCTTCTGTATACTAGTTTTCAGGAGAAGGTG 3120  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
3065 CTTAGAAATAGAAGATGTTCTAGTCAGAAATGCTCTTCTGTATACTAGTTTTCAGGAGAAGGTG 3124  
QY 3121 AAATTTGTAGAAGATCATCCAGGGAAACCTGTTTTCCCCTCTCGCATTTCTTGAAGATTC 3180  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
3125 AAATTTGTAGAAGATCATCCAGGGAAACCTGTTTTCCCCTCTCGCATTTCTTGAAGATTC 3184  
QY 3181 CCCGACTGAGAAGTGGAAAAATTTGGAAGGAGGAACATGACAAGATAATGATACGTGCT 3240  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
3185 CCCGACTGAGAAGTGGAAAAATTTGGAAGGAGGAACATGACAAGATAATGATACGTGCT 3244  
QY 3241 GTTTTAAAGCATGGGTACGGACGGTGGCAAGCTATTGTTGATGACAAAGAGTTGGGGATC 3300  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
3245 GTTTTAAAGCATGGGTACGGACGGTGGCAAGCTATTGTTGATGACAAAGAGTTGGGGATC 3304  
QY 3301 CAAGAGCTTATCTGCAAGAAATTTGAATTTCCCTCACATAAGTTTGTGCTGCTGAACAA 3360  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
3305 CAAGAGCTTATCTGCAAGAAATTTGAATTTCCCTCACATAAGTTTGTGCTGCTGAACAA 3364  
QY 3361 GCTGTTTGCAGGGGCAGAATGGTAGTGGGGCTCTAATCCGGGAGCACAGACTAACCCAG 3420  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
3365 GCTGTTTGCAGGGGCAGAATGGTAGTGGGGCTCTAATCCGGGAGCACAGACTAACCCAG 3424  
QY 3421 AATCCTGGAAGCGTTATTACTGGAACAATAATGCTTCTGCTGATGGGGCTCAAGTAAAC 3480  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
3425 AATCCTGGAAGCGTTATTACTGGAACAATAATGCTTCTGCTGATGGGGCTCAAGTAAAC 3484  
QY 3481 TCGATGTTCTATTATCGGGACATGCGAGAGACGACTTGTGTTAGTTTGTGAAAAAGCGAGTT 3540  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
3485 TCGATGTTCTATTATCGGGACATGCGAGAGACGACTTGTGTTAGTTTGTGAAAAAGCGAGTT 3544  
QY 3541 CTGCTTTTGGGAAGGCGATGAATTAATGAATACGAGAGGAATATTATGGACTTGGTGGC 3600  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
3545 CTGCTTTTGGGAAGGCGATGAATTAATGAATACGAGAGGAATATTATGGACTTGGTGGC 3604  
QY 3601 TCATCATCTATCCCTACTGAAGAACCCAGAAAGCTGAACCAAGATCGCTGACACAGTGGGA 3660

Db	3605	TCATCATCTATCCCTACTGAAGAACCCAGAAAGCTGAACCAAGATCGCTGACACAGTGGGA	3664
Qy	3661	GTGAGCTTTATTGAGGTTGATGATAAATGCTTGATGGACTTCCTAAGACTGATCCTATC	3720
Db	3665	GTGAGCTTTATTGAGGTTGATGATAAATGCTTGATGGACTTCCTAAGACTGATCCTATC	3724
Qy	3721	ACTTCAGAAAGAAATTATGGGGCTGCTGTTGACAACAACCAAGCGCGGGTCGAAATAGCT	3780
Db	3725	ACTTCAGAAAGAAATTATGGGGCTGCTGTTGACAACAACAACCAAGCGCGGGTCGAAATAGCT	3784
Qy	3781	CAACATTATAACCAAGATGTGCAAACTTCTTGATGAGAACGCTCGGGAATCAGTCCAAGCA	3840
Db	3785	CAACATTATAACCAAGATGTGCAAACTTCTTGATGAGAACGCTCGGGAATCAGTCCAAGCA	3844
Qy	3841	TATGTAAACAACCAACCAACCGAGTACCAAGGTGAATGAGAGCTTCCGTGCACCTCAAATCT	3900
Db	3845	TATGTAAACAACCAACCAACCGAGTACCAAGGTGAATGAGAGCTTCCGTGCACCTCAAATCT	3904
Qy	3901	ATCAATGGTAAACATTAAACACAATCCTTTCGATTACATCTGATCAATCCCAAGTCACATGAA	3960
Db	3905	ATCAATGGTAAACATTAAACACAATCCTTTCGATTACATCTGATCAATCCCAAGTCACATGAA	3964
Qy	3961	GACGACACCAAGCCAGACCTAAACAATGTTGAGATGAAGGACACGCGCCGAAGAAACAAA	4020
Db	3965	GACGACACCAAGCCAGACCTAAACAATGTTGAGATGAAGGACACGCGCCGAAGAAACAAA	4024
Qy	4021	CCGTTAAGAGGTGGCGTCGTCGATCTGAATGTGGTGGAGGGAGGAGGAACATTGCTGAA	4080
Db	4025	CCGTTAAGAGGTGGCGTCGTCGATCTGAATGTGGTGGAGGGAGGAGGAACATTGCTGAA	4084
Qy	4081	GCTAGTGGAAGTGTGATGTAAAAATGGAAGAACCCAAAGAAAGAGAGAGCCAAAGAAC	4140
Db	4085	GCTAGTGGAAGTGTGATGTAAAAATGGAAGAACCCAAAGAAAGAGAGAGCCAAAGAAC	4144
Qy	4141	ATGGTCGTTGATTGACTCAACTGGTAAATCAAGATTC	4177
Db	4145	ATGGTCGTTGATTGACTCAACTGGTAAATCAAGATTC	4181

RESULT 2	AF185577	4180 bp	mRNA	linear	PLN 30-NOV-1999
LOCUS	Arabidopsis thaliana	4180 bp	mRNA	linear	PLN 30-NOV-1999
DEFINITION	Arabidopsis thaliana chromatin remodeling factor CHD3 (PKL) mRNA, complete cds.				
ACCESSION	AF185577				
VERSION	AF185577.1	GI:6478517			
KEYWORDS					
SOURCE	Arabidopsis thaliana (thale cress)				
ORGANISM	Arabidopsis thaliana				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
REFERENCE	1 (bases 1 to 4180)				
AUTHORS	Ogas,J., Kaufmann,S., Henderson,J. and Somerville,C.				
TITLE	PICKLE is a CHD3 chromatin-remodeling factor that regulates the transition from embryonic to vegetative development in Arabidopsis				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 96 (24), 13839-13844 (1999)				
MEDLINE	20040637				
PUBMED	10570159				
REFERENCE	2 (bases 1 to 4180)				
AUTHORS	Ogas,J., Kaufmann,S., Henderson,J. and Somerville,C.				
TITLE	Direct Submission				
JOURNAL	Submitted (12-SEP-1999) Biochemistry, Purdue, 1153 Biochemistry, West Lafayette, IN 47907, USA				
FEATURES	Location/Qualifiers				
source	1. .4180				
	/organism="Arabidopsis thaliana"				
	/mol_type="mRNA"				
	/cultivar="Columbia"				
	/db_xref="taxon:3702"				
	/chromosome="2"				
	/map="between mi238 and GPA1"				

gene	1. .4180				
CDS	/gene="PKL"				
	25. .4179				
	/gene="PKL"				
	/note="PICKLE"				
	/codon_start=1				
	/product="chromatin remodeling factor CHD3"				
	/protein_id="AAF13875.1"				
	/db_xref="GI:6478518"				
	/translation="MSSLVERLRIRSDRKPVYNLDDSDDDFVPKKDRTFEQVEAIVR				
	TDAKENACQGESNLSVNCNTCTYAFHAKCLVPLPKDASVENWRCPCEVSPLEIDK				
	ILDCEMRPTKSEQSSDAEPKPIFVKQYLKWKGLSYLHCSWVPEKEFKAYKSNHR				
	LKTRVNNFHRQMESFNNSDDFVAIRPEWTTVDRIILACREEDGELEYLVKYKELSYDE				
	CYWESESDISTFQNEIQRFKDVNSRTRRSKVDHKNRPDRFQQFDHTPEFLKGLLHPY				
	QLEGLNPLRFSWSKQTHVILADEMGLGKTIQSIALLASLFEENLIPLHVIAPLSTLRN				
	WEREFATWAPQMNVMYFGTAQARAVIREHEFYLSKDQKKIKKKSGQISSESKQKRI				
	KFDVLLTSYEMINLDSAVLKPIKWECEMIVDEGHLRKNKDSKLFSSLTQYSSNHRILLT				
	GTPLNQNLDELFMLHFLDAGKFGSLEEFQEEFKDINOBEQISRLHKMLAPHLRRVK				
	KDVMKMPKKELILRVDLSSLOKEYYKAIFTRNYQVLTKKGAQISLNNIMMELRKV				
	CCHPYMLEGVEPVIHDANEAPKQLLESCKQLQLLDKMMVKLKEQGHVLIYTFQFHML				
	DLLEDYCTHKWQYERIDGKVGAERQIRIDRFNAKNKFCFLLSSTRAGGLINLAT				
	ADTVIYDSDNPHADLQAMARAHRLGQTNKVMYIRLINRGTIIEERMQLTKKMWLE				
	HLVVGKLKTNINQEEDDIIRYGSKELFASEDDEAGSKGIHYDDAAIDKLLDRDLV				
	EAEFVSVDDEEENGFLKAFKANFEYIDENEAALAEQVAAESKSSAGNSDRASYWE				
	ELLKDKFELHQAEEELNALGKRKSRKQLVSI EDDDLGLEDVSSDGESEYEAESTDGE				
	AACQGVQTRRPPYRRKGRDNL EPTPLMEGEGRSFRLVGNFSQRAIFVQTLRMRYGAGN				
	FDWKEFVRLKQKTFEIN EYILFLKHIAEEDISNPFSFDGVPKEGLRIEDVLRI				
	ALLILVQEKVFEVDPGKVPFSPRILERFPLRSRSGKIWEHDKIMIRAVLKHGYGR				
	WQAIVDKDELQIOLICKELNFPHISLSAAEQAGLQNGSGSNPQAQYNTQNPQSVI				
	TGNNNASADGAQVNSMFFYRDMQRRLLVEFVKRVLLEKAMNVEYEAAYEYGLGSSSI				
	PTEEPEAEPKIADTVGVSFIEVDDMLDGLPKTDPITSEIMGAADVNNQARVEIAQH				
	YNQCKLLDENARESQAIVNNQPPSTKVNESFRALKSINGNINTILSITSDQSKSHE				
	DDTKPDLNNVEMKDTAEETKPLRGVVVDLNVVEGEENIAPASGSDVVKMEAEKEEK				
	KNMVVD"				

ORIGIN

Query Match	99.5%;	Score 4156;	DB 8;	Length 4180;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 4156;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	ATGAGTAGTTTGGTGGAGAGGCTTCGCATACGATCTGATAGGAAACAGTTTATAACCTA	60	
Db	25	ATGAGTAGTTTGGTGGAGAGGCTTCGCATACGATCTGATAGGAAACAGTTTATAACCTA	84	
Qy	61	GATGATTCGATGATGACGACTTCGTTCTCTAAAAAAGATCGAACCTTTGAGCAAGTCGAG	120	
Db	85	GATGATTCGATGATGACGACTTCGTTCTCTAAAAAAGATCGAACCTTTGAGCAAGTCGAG	144	
Qy	121	GCTATTGTCAGAACTGATCGAAAGAAAATGTCATGTGAGGCTTGTGGGAAAAGTACTAAT	180	
Db	145	GCTATTGTCAGAACTGATCGAAAGAAAATGTCATGTGAGGCTTGTGGGAAAAGTACTAAT	204	
Qy	181	CTTGTAAGCTGCAATACATGCACCTTATGCGTTCCTCATGCTAAATGCTTAGTTCACCTCTT	240	
Db	205	CTTGTAAGCTGCAATACATGCACCTTATGCGTTCCTCATGCTAAATGCTTAGTTCACCTCTT	264	
Qy	241	AAAGATGCTTCGTTGGGAAAATTGGAGATGCCCTGAATGTGTTAGTCTCTTAACGAGATA	300	
Db	265	AAAGATGCTTCGTTGGGAAAATTGGAGATGCCCTGAATGTGTTAGTCTCTTAACGAGATA	324	
Qy	301	GATAAGATATTGGATTGTGAAATGCGTCTCTACAAAATCTAGTGAAACAGGTTCCCTCCGAT	360	
Db	325	GATAAGATATTGGATTGTGAAATGCGTCTCTACAAAATCTAGTGAAACAGGTTCCCTCCGAT	384	
Qy	361	GCGGAACCGAAGCCAATTTTGTGAAACAGATATCTCGTGAAGTGGAAGGATTATCATAC	420	
Db	385	GCGGAACCGAAGCCAATTTTGTGAAACAGATATCTCGTGAAGTGGAAGGATTATCATAC	444	
Qy	421	CTTCACCTGCTCTTGGGTGCCTGAGAGGAGTTCAGAAAGCTTATAAGTCAAATCATCGT	480	
Db	445	CTTCACCTGCTCTTGGGTGCCTGAGAGGAGTTCAGAAAGCTTATAAGTCAAATCATCGT	504	
Qy	481	TTAAAAACCCAGAGTGAACAATTTTCCACCGTCAAAATGGAGTCAATTCATAAACAGCGAAGAT	540	



Db	505	TTAAAAAC	CAGAGTGA	ACAAATTT	TCACCGT	CAATGGAG	TCAATCA	ATAACG	CAAGAT	564															
QY	541	GATTTTGT	TGCCATA	CGTCCT	GAGTGG	ACCAC	TGTTGAT	CGGATT	CTTTGCC	TGCAGAGAG	600														
Db	565	GATTTTGT	TGCCATA	CGTCCT	GAGTGG	ACCAC	TGTTGAT	CGGATT	CTTTGCC	TGCAGAGAG	624														
QY	601	GAAAGAT	GGGAGCT	GGAATAT	CTTGT	CAAAAT	ATAAAG	AGCTAT	CCTAT	GATGAATGTTAT	660														
Db	625	GAAAGAT	GGGAGCT	GGAATAT	CTTGT	CAAAAT	ATAAAG	AGCTAT	CCTAT	GATGAATGTTAT	684														
QY	661	TGGGAGT	CAGAATC	AGACAT	CTCA	ACCTT	CCAGAA	TGAAAT	TTCAA	AGGTTCAAGGATGTA	720														
Db	685	TGGGAGT	CAGAATC	AGACAT	CTCA	ACCTT	CCAGAA	TGAAAT	TTCAA	AGGTTCAAGGATGTA	744														
QY	721	AATTCTA	GAACCT	CGCAGA	AGTAA	AGATG	TGTG	ACCATA	AAAGAA	ATCCCAGAGACTTTCAA	780														
Db	745	AATTCTA	GAACCT	CGCAGA	AGTAA	AGATG	TGTG	ACCATA	AAAGAA	ATCCCAGAGACTTTCAA	804														
QY	781	CAGTTTG	ATCAVAC	TCTCT	GAATTC	CTCAA	AGGCTTG	TTACAT	CCATAC	CCAGCTTGAGGGA	840														
Db	805	CAGTTTG	ATCAVAC	TCTCT	GAATTC	CTCAA	AGGCTTG	TTACAT	CCATAC	CCAGCTTGAGGGA	864														
QY	841	CTTAAT	TTTTT	TGCGGT	TCTCG	TGTC	CAAA	ACAG	ACGCA	TGTAATCCTTTGCTGATGAAATG	900														
Db	865	CTTAAT	TTTTT	TGCGGT	TCTCG	TGTC	CAAA	ACAG	ACGCA	TGTAATCCTTTGCTGATGAAATG	924														
QY	901	GGACTAG	GCAAG	CAATTC	AAAGCA	TGCCC	TTTAG	CTTCA	CTTTT	TGAGGAGAACCTC	960														
Db	925	GGACTAG	GCAAG	CAATTC	AAAGCA	TGCCC	TTTAG	CTTCA	CTTTT	TGAGGAGAACCTC	984														
QY	961	ATTCG	CAATTTG	GTAATTG	CTC	TATC	GACTCT	GCGTAA	CTGGG	AGAGAGAGTTTGCC	1020														
Db	985	ATTCG	CAATTTG	GTAATTG	CTC	TATC	GACTCT	GCGTAA	CTGGG	AGAGAGAGTTTGCC	1044														
QY	1021	ACATGG	CCCCC	ACAGAT	GAAC	GTG	GTAT	TGTAT	TTTGG	CAC	TGCGCAAGCTCGAGCAGTT	1080													
Db	1045	ACATGG	CCCCC	ACAGAT	GAAC	GTG	GTAT	TGTAT	TTTGG	CAC	TGCGCAAGCTCGAGCAGTT	1104													
QY	1081	ATCAG	AGAAC	ATGAG	TTTACT	TATC	GAAAG	ATCA	AAAAA	AGATCA	AGAAAAAGAAATCT	1140													
Db	1105	ATCAG	AGAAC	ATGAG	TTTACT	TATC	GAAAG	ATCA	AAAAA	AGATCA	AGAAAAAGAAATCT	1164													
QY	1141	GGACAAA	ATAAG	TAGG	CAAA	GCAAG	CAAAAA	AAGAA	TCAAG	TTTGAT	GTCTCCTCACATCG	1200													
Db	1165	GGACAAA	ATAAG	TAGG	CAAA	GCAAG	CAAAAA	AAGAA	TCAAG	TTTGAT	GTCTCCTCACATCG	1224													
QY	1201	TATGAG	ATGAT	CAAC	CTAG	ATTCA	GAGATT	CTAAAA	CCAAT	TAAAG	TGGAGTGCATGATT	1260													
Db	1225	TATGAG	ATGAT	CAAC	CTAG	ATTCA	GAGATT	CTAAAA	CCAAT	TAAAG	TGGAGTGCATGATT	1284													
QY	1261	GTTGAT	GAA	GGTCA	TGACT	GAAAA	ATAAG	GATTCA	AAAGCT	GTTCT	CTTCATTGACACAG	1320													
Db	1285	GTTGAT	GAA	GGTCA	TGACT	GAAAA	ATAAG	GATTCA	AAAGCT	GTTCT	CTTCATTGACACAG	1344													
QY	1321	TATTC	AAAGTA	ACC	CGTAT	TCTT	CTGAC	AGGA	ACAC	CACTT	CAGAA	CAACTTGGATGAA	1380												
Db	1345	TATTC	AAAGTA	ACC	CGTAT	TCTT	CTGAC	AGGA	ACAC	CACTT	CAGAA	CAACTTGGATGAA	1404												
QY	1381	CTTTT	TCATG	CTATG	CAAT	TTTCT	TGATG	CGGGG	AAAGTT	TGGA	AGTTTGG	AGGAGTTCCAG	1440												
Db	1405	CTTTT	TCATG	CTATG	CAAT	TTTCT	TGATG	CGGGG	AAAGTT	TGGA	AGTTTGG	AGGAGTTCCAG	1464												
QY	1441	GAGG	AGTTCAA	AGATAT	TAA	TCAAG	AGGAG	CA	GATCT	CAAG	GTTC	CAAAA	TGTTGGCT	1500											
Db	1465	GAGG	AGTTCAA	AGATAT	TAA	TCAAG	AGGAG	CA	GATCT	CAAG	GTTC	CAAAA	TGTTGGCT	1524											
QY	1501	CCAC	ATTTG	CTC	AGA	AGGG	TAAAA	AAAA	AGAC	CGTA	TATG	AAAG	ACATGCC	CCCC	AAAA	AGGAG	1560								
Db	1525	CCAC	ATTTG	CTC	AGA	AGGG	TAAAA	AAAA	AGAC	CGTA	TATG	AAAG	ACATGCC	CCCC	AAAA	AGGAG	1584								
QY	1561	CTCAT	TTT	TG	CGT	GTTG	ATCT	G	AGC	AGT	CTG	C	AG	AA	GA	ATAT	TAC	AA	AG	CT	AT	TTT	T	ACC	1620



QY	2701	GATGGAGATGAAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGACAAGGAGTT	2760
Db	2725	GATGGAGATGAAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGACAAGGAGTT	2784
QY	2761	CAGACGGGTCGACGGCCGTA CAGAAGAAAGGTCGCGATAAATTTGGAAACCAACTCCGTTG	2820
Db	2785	CAGACGGGTCGACGGCCGTA CAGAAGAAAGGTCGCGATAAATTTGGAAACCAACTCCGTTG	2844
QY	2821	ATGGAAGGTGAGGGAGATCTTT CAGAGTACTGGGTTTCAACCAGAGTCAAAAGGCCATT	2880
Db	2845	ATGGAAGGTGAGGGAGATCTTT CAGAGTACTGGGTTTCAACCAGAGTCAAAAGGCCATT	2904
QY	2881	TTTGTACAGACTTTGATGAGGTATGGAGCTGGCAATTTTGATTGGAAGGAGTTTGTTCCT	2940
Db	2905	TTTGTACAGACTTTGATGAGGTATGGAGCTGGCAATTTTGATTGGAAGGAGTTTGTTCCT	2964
QY	2941	CGCTTAAAGCAGAAACCTTTTGAAGAAATAAATGAATATGGAATACTCTTCTTGAAGCAC	3000
Db	2965	CGCTTAAAGCAGAAACCTTTTGAAGAAATAAATGAATATGGAATACTCTTCTTGAAGCAC	3024
QY	3001	ATTGCTGAAGAAATAGACGAGAATTCTCCAACCTTTT CAGATGGTGTGCCCAAGGAAGGA	3060
Db	3025	ATTGCTGAAGAAATAGACGAGAATTCTCCAACCTTTT CAGATGGTGTGCCCAAGGAAGGA	3084
QY	3061	CTTAGAATAGAAGATGTTCTAGTCAGAAATTGCTCTTCTGATACTAGTTTCAGGAGAAAGTG	3120
Db	3085	CTTAGAATAGAAGATGTTCTAGTCAGAAATTGCTCTTCTGATACTAGTTTCAGGAGAAAGTG	3144
QY	3121	AAATTTGTAGAAGATCATCCAGGGAAA CCTGTTCCTCTCGCATTCCTTGAAGATTCT	3180
Db	3145	AAATTTGTAGAAGATCATCCAGGGAAA CCTGTTCCTCTCGCATTCCTTGAAGATTCT	3204
QY	3181	CCCGGACTGAGAAAGTGGAAGAAATTTGGAAGGAGGAACATGACAAGATAATGATACGTGCT	3240
Db	3205	CCCGGACTGAGAAAGTGGAAGAAATTTGGAAGGAGGAACATGACAAGATAATGATACGTGCT	3264
QY	3241	GTTTTAAAGCATGGGTACGACGGTGGCAAGCTATTGTTGATGACAAAGAGTTGSGGATC	3300
Db	3265	GTTTTAAAGCATGGGTACGACGGTGGCAAGCTATTGTTGATGACAAAGAGTTGSGGATC	3324
QY	3301	CAAGAGCTTATCTGCAAGAAATTGAAATTTCCCTCACATAAGTTTGTCTGCTGCTGAACAA	3360
Db	3325	CAAGAGCTTATCTGCAAGAAATTGAAATTTCCCTCACATAAGTTTGTCTGCTGCTGAACAA	3384
QY	3361	GCTGTTTGCAGGGCAGAAATGGTAGTGGGGCTCTAATCCGGGAGCACAGACTAACCCAG	3420
Db	3385	GCTGTTTGCAGGGCAGAAATGGTAGTGGGGCTCTAATCCGGGAGCACAGACTAACCCAG	3444
QY	3421	AATCCTGGAAGCGTTATTACTGGGAA CAATAATGCTTCTGCTGATGGGCTCAAGTAAAC	3480
Db	3445	AATCCTGGAAGCGTTATTACTGGGAA CAATAATGCTTCTGCTGATGGGCTCAAGTAAAC	3504
QY	3481	TCGATGTTCTATTATCGGGACATGCAGAGACGACTTGTGTAGTTTGTGAAAAAGCGAGTT	3540
Db	3505	TCGATGTTCTATTATCGGGACATGCAGAGACGACTTGTGTAGTTTGTGAAAAAGCGAGTT	3564
QY	3541	CTGCTTTTGGAGAAGGCGATGAATTATGAATACGCAGAGGAATATTATGGACTTTGGTGGC	3600
Db	3565	CTGCTTTTGGAGAAGGCGATGAATTATGAATACGCAGAGGAATATTATGGACTTTGGTGGC	3624
QY	3601	TCATCATCTATCCCTACTGAAGAAC CAGAGCTGACCAAAAGATCGCTGACACAGTGGGA	3660
Db	3625	TCATCATCTATCCCTACTGAAGAAC CAGAGCTGACCAAAAGATCGCTGACACAGTGGGA	3684
QY	3661	GTGAGCTTTATTGAGGTTGATGAAATGCTTGTATGGACTTCCTAAGACTGATCCTATC	3720
Db	3685	GTGAGCTTTATTGAGGTTGATGAAATGCTTGTATGGACTTCCTAAGACTGATCCTATC	3744
QY	3721	ACTTCAGAAGAAATTA TGGGGCTGCTGTTGACAA CCAAGCGGGT CGAAATAGCT	3780
Db	3745	ACTTCAGAAGAAATTA TGGGGCTGCTGTTGACAA CCAAGCGGGT CGAAATAGCT	3804

Qy	3781	CAACATTATTAACCCAGATGTGCAAACTTCTTGATGAGAACCGCTCGGGAAATCAGTCCAAGCA	3840
Db	3805		
Db	3805	CAACATTATTAACCCAGATGTGCAAACTTCTTGATGAGAACCGCTCGGGAAATCAGTCCAAGCA	3864
Qy	3841	TATGTAAAACAACCACCCGAGTACCAAAGTGAAATGAGAGCTTCCTGTCACCTCAAAATCT	3900
Db	3865		
Db	3865	TATGTAAAACAACCACCCGAGTACCAAAGTGAAATGAGAGCTTCCTGTCACCTCAAAATCT	3924
Qy	3901	ATCAATGGTTAACAATTAAACA CAATCCTTTTCGATTACATCTGATCAATCCAAGTCACATGAA	3960
Db	3925		
Db	3925	ATCAATGGTTAACAATTAAACA CAATCCTTTTCGATTACATCTGATCAATCCAAGTCACATGAA	3984
Qy	3961	GACGACACCAAGCCAGACCTTAAACAATGTTTGATGAAGGACACCGGCCGAAGAAAACAAAA	4020
Db	3985		
Db	3985	GACGACACCAAGCCAGACCTTAAACAATGTTTGATGAAGGACACCGGCCGAAGAAAACAAAA	4044
Qy	4021	CCGTTAAGAGGTGGCGTCTGTCGATCTGAATGTGGTGGAGGGAGAGAACATTGCTGAA	4080
Db	4045		
Db	4045	CCGTTAAGAGGTGGCGTCTGTCGATCTGAATGTGGTGGAGGGAGAGAACATTGCTGAA	4104
Qy	4081	GCTAGTGGAAAGTGTGTTGATGTAAAAATGGAAGAAGCCAAAGAAAGAGAGCCAAAGAAC	4140
Db	4105		
Db	4105	GCTAGTGGAAAGTGTGTTGATGTAAAAATGGAAGAAGCCAAAGAAAGAGAGCCAAAGAAC	4164
Qy	4141	ATGCTCGTTGATTGAC	4156
Db	4165		
Db	4165	ATGCTCGTTGATTGAC	4180

```

RESULT 3
AX507831
LOCUS          AX507831          3204 bp      DNA
DEFINITION    Sequence 2526 from Patent WO0216655.
ACCESSION     AX507831
VERSION       AX507831.1  GI:23389068
KEYWORDS
SOURCE        Arabidopsis thaliana (thale cress)
ORGANISM      Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1
AUTHORS       Harper, J.F., Kreps, J., Wang, X. and Zhu, T.
TITLE         Stress-regulated genes of plants, transgenic plants containing
              same, and methods of use
JOURNAL       Patent: WO 0216655-A 2526 28-FEB-2002;
              The Scripps Research Institute (US) ; Syngenta Participations AG
              (CH)
FEATURES
            source      Location/Qualifiers
            1..3204
                /organism="Arabidopsis thaliana"
                /mol_type="unassigned DNA"
                /db_xref="taxon:3702"
ORIGIN

```

[illegible]

Db 301 AAGGCTTACAAGTCCCATCCTCATTTTAAAACTAAAAATAAGGGTACACAGGTTTAAACGCT 360

Qy 511 CAAATGGAGTCATTCA-----ATAACAGCGAAGATGATTTTGTGGCCATACGTCCTGAG 564

Db 361 GCTATGGACGTAATTCATCGCTGAGAACGGAGCTCATGAGTTTATCGCCATTGTCCTGAA 420

Qy 565 TGGACCACTGTTGATCGGATCTTGCCTGCAGAGAGGAAGATGGGAGCTGGAATATCTT 624

Db 421 TGGAAAACTGTTGACCGGATTATTGCTTGC CGGAAGGAGATGATGGGAGGATATCTG 480

Qy 625 GTCAAATATAAAGAGCTATCCTATGATGAATGTTATTGGGAGTCAGAATCAGACATCTCA 684

Db 481 GTGAAGTACAAAGAACTTCTTACAGAAATAGTTACTGGGAATCTGAATCTGACATCTCA 540

Qy 685 ACCTTCCAGAAATGAAATTCAAAGGTTCAAGGATGTAAATTTCTAGAACTCGCAGAAGTAAA 744

Db 541 GATTTCCAGAAATGAAATCCAAAGGTTTAAAGGATATAAACTTAGCTCTCGTAGAGATAAA 600

Qy 745 GATGTTGACCAATAAAGAAATCCCAGAGACTTTCAACAGATTGATCATFACTCCTGAATTC 804

Db 601 TATGTCGAAAAATGAGAGAAATCGAGAGGAATTCAAAACAGTTTGATCTTACTCCTGAATTT 660

Qy 805 CTCAAAGGCTTGTATACATCCATACCACTTGAGGACTTAATTTTTCGGTCTCGTGG 864

Db 661 CTTACAGGCACATTGCATACATACCAGCTCGAAGGCTGAATTTCTTAAGGTATTCTATGG 720

Qy 865 TCAAAACAGACGCGATGTAATCCTTGCTGATGAAATGGGACTAGGCAAGACAAATTCAAAGC 924

Db 721 TCCAAAGAAAACCAATGTAACTCTTGGGATGAAATGGGACTAGGCAAAACTATTTCAGAGC 780

Qy 925 ATTGCCCTTTTAGCTTCACTTTTGTGAGGAGAACCTCATTCCTCGCATTTGGTAATTCCTCT 984

Db 781 ATTGCCCTTCTTGGCCTCTTTGTTTCGAGGAAAACTTATCTCCTCATCTAGTTGTTGCTCCT 840

Qy 985 CTATCGACTCTGGTAACTGGGAGAGAGATTGTCACATGGGCCCCACAGATGAACGTG 1044

Db 841 CTCTCAACCATCCGTAACCTGGGAGAGGGAATTTGCCACCTGGGCACCACATATGAATGTG 900

Qy 1045 GTTATGTATTTTGGCACTGCGCAAGCTCGAGCAGTTATCAGAGAACATGAGTTTACTTA 1104

Db 901 GTTATGTACACTGGGATTCGGAAGCACGTCGATGTTATATGGAACATGAGTTCTACTTC 960

Qy 1105 TCGAAAGATCAAAAAAGATCAAGAAAAAGAAATCTGGACAAATAAGTAGCGAAAGCAAG 1164

Db 961 TCTGAAGGGCGTAAAAGGTTAAAG----- 984

Qy 1165 CAAAAAAGATCAAGTTTGATGTCCTCTCATCGTATGAGATGATCAACCTAGATTCA 1224

Db 985 ----- 984

Qy 1225 GCAGTTCTAAAAACCAATTAAGTGGGAGTGCATGATTGTTGATGAAGTCAATCGACTGAAA 1284

Db 985 ----- 984

Qy 1285 AATAAGGATTCAAAGCTGTTCTCTTCATTGACACAGTATTCAAGTAACCAACCGTATTCTT 1344

Db 985 ----- 984

Qy 1345 CTGACAGGAACACCACCTTCAGAACAACTTGATGAACCTTTTTCATGCTCATGCTTTTCTT 1404

Db 985 ----- 984

Qy 1405 GATCGGGGAAGTTTGGAAAGTTTGGAGGAGTTCCAGGAGGAGTTCAAGATATTAATCAA 1464

Db 985 ----- 984

Qy 1465 GAGGAGCAGATCTCAAGGTTGCACAAAATGTTGGCTCCACATTTGCTCAGAAGGGTAAAA 1524

Db 985 -----AAA 987

Qy 1525 AAAGACGTAATGAAGACATGCCCCCCCAAAAAGGAGCTCATTTTTCGCTGTTGATCTGAGC 1584

Db 988 GACGTGCTCAAGGATAAGGTGCCCCCAAGAGGAGCTCATTTTTCGCTGTTGATATGAGC 1047

Qy 1585 AGTCTGCAGAAAGAAATATTACAAAGCTATTTTACCCTAATATTATCAAGTATTTGACAAAA 1644

Db 1048 AGCCAGCAAAAAGAAAGTGTACAAAGCTGTTATTACCATAATATTATCAAGTCTTGACAAAG 1107

Qy 1645 AAGGGAGGTGCTCAAATTTCCCTTAATAAACATTTATGATGGAATTTAGGAAAAAGTATGCTGC 1704

Db 1108 AAACGAGATGCTAAAAATATC-----TAACTGTCTAATGAAATTTAGCCCAAGTTTGCTCT 1161

Qy 1705 CATCTTATATGCTAGAGGGTGTGAGCCAGTTTACAGCCAGTTATTACAGACGCAATGAAGCTTTCAAA 1764

Db 1162 CATCCGTATCTGTTACCAGACTTTGAGCCCAAGATTGGAAGACGCAATGAAGCTTTTACA 1221

Qy 1765 CAACTTTTGGAGTCTTGTGGAAAGCTGCAACTTCTAGATAAAATGATGGTCAAACTGAAA 1824

Db 1222 AAACATATTGGAAGCTTCCGGAAAAGTTGCAACTGTTAGACAAAATGATGGTCAAACTCAA 1281

Qy 1825 GAGCAAGGACACAGAGTCCCTAATATATACACACAGTTTTCAGCATATGCTGGACTTACTTGAA 1884

Db 1282 GAGCAGGGCACAGAGTGTAAATCTATATACCCAAATCCAACACACCTTTTACCTTTTGGAA 1341

Qy 1885 GACTACTGTACCCATAAGAAAATGGCAGTACGAGCGAAATTTGATGGAAAAGTTGGCGGAGCT 1944

Db 1342 GACTACTTCACTTTCAAGAATTGGAATTTATGAACGGATTGATGGAAAAGATCAGTGGACCT 1401

Qy 1945 GAGCGCAAAATACGCATAGATCGGTTCAATGCCAAAAATTTCTAACAAAGTTTGTGTTTTTG 2004

Db 1402 GAGAGGCAAGTAAGAATAGATCGGTTTAAATGCAGAAAACTCGAACCGGTTCTGTGTTTTTG 1461

Qy 2005 CTCTCCCAAGAGCTGGTGGCTTAGGAATAAATCTTGCAACGGCTGATACAGTAATCAT 2064

Db 1462 CTCAGTACAAGAGCTGGTGGAAATAGGATAAACCTTGCAACAGCTGATACAGTCATCAT 1521

Qy 2065 TATGACAGTACTGGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATCGACTT 2124

Db 1522 TATGACAGTATTGGAATCCTCATGCTGATCTCCAGGCTATGGCTAGAGTTTCATCGGCTT 1581

Qy 2125 GGCCAAACAAATAAGGTGATGATTTATATAGGCTCATAAACCGAGGCCATTTGAAGAAAG 2184

Db 1582 GGCCAAACAAATAAGGTAATGATCTACAGACTGATTTCACAAAGGTACTGTAGAAGAACGG 1641

Qy 2185 ATGATGCAATTGACTAAAAAGAAAATGGTTCTAGAGCATCTTGTGTTGGGAAAACCTCAA 2244

Db 1642 ATGATGGAGATAACCAAGAACAAAGATGCTTCTAGAGCATCTCGTCTGCTGGG-----A 1692

Qy 2245 ACACAAAAACATTAATCAGGAAGAGTTAGATGACATCATCAGGTATGATCAAAAGGAGCTT 2304

Db 1693 AAACAACATCTTTGTCAGGACGAGTTAGATGACATCATCAAAATGTTGTTCCAAGGAACTT 1752

Qy 2305 TTTGCTAGTGAAGATGATGAAGCAGGAAAAGTCTGGAAAAATTCATTATGATGATCGGCT 2364

Db 1753 TTCTCTGAAGAGAATGATGAAGCAGGAGGCTCTGGAAAAATCCATTACGATGATGCTGCT 1812

Qy 2365 ATAGACAAATTTGCTTGTATCGTATCTCGTGGAGGCAGAGGAGTCTCAGTGGATGATGAA 2424

Db 1813 ATCGAACAGTTGCTAGATCGTAATCATGTAGATGCTGTGGAGGTTCTACTGGATGATGAG 1872

Qy 2425 GAGGAGATGGATTCTTAAAGGCTTCAAGGTGGCTAAATTTTGAATATATAGATGAAAAAT 2484

Db 1873 GAGGAAACTGATTTTTTAAAGAAATTTCAAGGTGGCTAGTTTTTGTAGTATGTAGATGATA 1932

Qy 2485 GAGGCACAGCATTAGAGGCGACAGAGATCGCTGCTGAAAGCAAAATCTTCAGCAGGCAAT 2544

Db 1933 AATGAGGCTGCAGCATTAGAGGAGGCACAAGCTATAGAAAACAATTTCTTCAGTAAGAAAT 1992

Qy 2545 TCTGATAGAGCAAGTTATTGGGAAGAGTTGTTAAAGATAAAATTTGAGCTGCACCAGGCT 2604

Db 1993 GCTGATAGAACAAAGTCACTGGAAAGATTGCTGAAAGACAAATATGAAGTGCAACAAGCT 2052

Qy 2605 GAGGAGCTTAATGCTCTTGGAAAAAGGAAGAGAAAGTCGCAAGCAGTTGGTATCCATTGAA 2664

Db 2053 GAGGAGCTCAGTCTCTTGGAAAGAGGAAGAGAAACGGCAAGCAGGTTGATGTA---TGCT 2109

QY 2665 GAAGATGATCTTGCTGGTTTGGAAAGATGTGAGCTCTGATGGAGATGAAAGTTATGAAGCT 2724  
|||||  
Db 2110 GAAGATGATCTTGATGGTTTAGAAGAAATCTCTGATGAGGAGGACGAATATGTCTTGAT 2169  
QY 2725 GAGTCAACA-----GATGGTGAAGCAGCAGGACAA 2754  
|||||  
Db 2170 GATTTAAAGTGACCTCTGATGAGGAGGAGGAGCAGATGAACCCGAAGCAGCTCGACAG 2229  
QY 2755 GGAGTTC---AGACGGGTCGACGGCCGTACAGAAGAAAGGTCGCGATAAATTGGAACCA 2811  
|||||  
Db 2230 AGAAAGCCGAGGACGGTTACAAGCCCTTACAGAAAAAGGGCTCGCGATAAATTCGGAGGAA 2289  
QY 2812 ACTCCGTTGATGGAAGGTGAGGGGAGATCTTTTCAGAGTACTGGGTTTCAACCAAGATCAA 2871  
|||||  
Db 2290 ATACCTTGTATGGAAGCGAAGGAGATATCTCATGGTGTGGATTAAACGAGACTGAG 2349  
QY 2872 AGGGCCATTTTGTACAGACTTTGATGAGGATGGAGCTGGCAATTTTTCATTGGAAGGAG 2931  
|||||  
Db 2350 AGGATATATTTTACGTACGTTTAAGAGGTATGGAATCTTATTTTGTGAACACATTGCC 2409  
QY 2932 TTTGTTCTCGCTTAAAGCAGAAGACCTTTTGAAGAAATAAATGAATATGGAATACTCTTC 2991  
|||||  
Db 2410 GAAAACCTTACCATAAATTTCTACAAATTTTAAAGTTATTACGGCAATGGT----- 2459  
QY 2992 TTGAAGCACATTGCTCAAGAAATAGACGAGAAATCTCCAACCTTTTTCAGATGGTGTGCCC 3051  
Db 2460 -----CTATGCAGATGGAGTTCCC 2478  
QY 3052 AAGGAAGGACTTAGAATAGAAGATGTTCTAGTCAGAAATGCTCTTCTGATACTAGTTTCAG 3111  
Db 2479 AAAGAAGGGATAAGCACTGACGAGCTACTAGTGAGCATGACTTTTATGATGCTAGTAAAG 2538  
QY 3112 GAGAAGGTGAAATTTGTAGAAGATCATCCAGGGAACCTGTTTTCCTCCTCGCATTCCT 3171  
Db 2539 GAGAAGTGTCAATTTTGGACAACCATCCGACAGCACCTGTTTT---CTCTAACTACGTC 2595  
QY 3172 GAAAGATTTCCCGGACTGAGAAAGTGGAAAATTTGGAAGGAGGAACATGACAAGATAATG 3231  
Db 2596 ATCAGTAAGTACAATTTGAGAAATGGAGCATTTTCTAAGGAAGAACATGATAGGATACTG 2655  
QY 3232 ATACGTGCTGTTTTAAAGCATGGGTACGGACGGTGGCAAGCTATTGTTGATGACAAAGAG 3291  
Db 2656 ATTCCTGCTGTTTCCAAGCATGGCTATGTTAGATGGGTGGCCATCGTTGAAGACGAAGAG 2715  
QY 3292 TTGGGGATCCAAGAGCTTATCTGCAAAGAAATTGAATTTCCCT 3333  
Db 2716 ATAGGGTTCCAAGAGGTTGCCTGCAAGACCTGTAATATCCCT 2757

RESULT 4  
BC046866  
LOCUS  
DEFINITION  
Xenopus laevis chromodomain helicase DNA binding protein 4, mRNA  
(cdna clone MGC:52739 IMAGE:4683903), complete cds.  
ACCESSION  
BC046866  
VERSION  
BC046866.1 GI:28422179  
KEYWORDS  
MGC.  
SOURCE  
Xenopus laevis (African clawed frog)  
ORGANISM  
Xenopus laevis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
Xenopodinae; Xenopus.  
REFERENCE  
1 (bases 1 to 6398)  
Klein,S.L., Strausberg,R.L., Wagner,L., Pontius,J., Clifton,S.W.  
and Richardson,P.  
TITLE  
Genetic and genomic tools for Xenopus research: The NIH Xenopus  
initiative  
JOURNAL  
Dev. Dyn. 225 (4), 384-391 (2002)  
MEDLINE  
22341132  
PUBMED  
12454917  
REFERENCE  
2 (bases 1 to 6398)  
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,  
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,

Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,  
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,  
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,  
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,  
Scheetz,T.B., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,  
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,  
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,  
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,  
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,  
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,  
Fahey,J., Helton,E., Kettman,M., Madan,A., Rodrigues,S.,  
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,  
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,  
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,  
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,  
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
22388257  
12477932  
3 (bases 1 to 6398)  
Klein,S. and Strausberg,R.  
Direct Submission  
Submitted (14-FEB-2003) National Institutes of Health, Xenopus Gene  
Collection (XGC), National Institute of Child Health and Human  
Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD  
20892-7510, USA  
NIH-MGC Project  
Contact: XGC help desk  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Dr. Igor Dawid  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Institute for Systems Biology  
http://www.systemsbiology.org  
contact: anadan@systemsbiology.org  
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettman, Anuradha  
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
Series: IRAX Plate: 94 Row: g Column: 7  
This clone was selected for full length sequencing because it  
passed the following selection criteria: Hexamer frequency ORF  
analysis, Similarity but not identity to protein.  
Location/Qualifiers  
1. .6398  
/organism="Xenopus laevis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8355"  
/clone="MGC:52739 IMAGE:4683903"  
/tissue\_type="Embryo, stage 31/32, Xenopus"  
/clone\_lib="NICHD\_XGC\_Emb4"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
1. .6398  
/gene="b230399n07-prov"  
/note="synonym: MGC52739"  
/db\_xref="LocusID:380196"  
178. .5859  
/codon\_start=1  
/product="b230399n07-prov protein"  
/protein\_id="AAH46866.1"  
/db\_xref="GI:28422180"  
/db\_xref="LocusID:380196"  
/translation="MASGLGPSGSDGEIEVMLNSSIHQDDDLNEDLSEVETPKIKK  
KKPKKAKESKSTKVKRQKEDVDDSSGENDYGDNDGGPIRSDSEGSDYTPGKKKK  
KMTPKREKSKVKRDEDDDDDDSKPKSSGQLDDWGMEIDHVFTEDDYRTLT  
NYKAFSQFVRPLIAAKNPKIAVSKMMVLGAKWREFSTNNPLKSSGSASVAAAAAV  
AVESMVASTEIVPPPPPPPPVVEVPLRKAKTKGKGNARKKPKVVRPPETKKT  
KAKKVAPLKIKLGGFGSKRKRSSSEEDIDADSDYDEGSMNSLSVSDGSTSKSSRSRK  
KLKATKKKKKEEGAAAGADGYETHQDYCEVCQQGGEIILCDTCPRAYHMVCLDPDM

TITLE

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

REMARK  
COMMENT

FEATURES  
source

gene

CDS





Db 3515 TCCTCCTGCTACTCGTCTGGAGGTTTAGGAATTAATCTGGCAACTGCAGATACTGTAG 3574

QY 2060 TCATTTATGACAGTGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC 2119

Db 3575 TCATATACGACTCGGACTGGAATCCTCACAATGATATTAGGCTTTACAGCGGCTCATC 3634

QY 2120 GACTTGCCCAACAAATAAGGTGATGATTTATAGGCTCATAAACCGAGGCACCATTTGAAG 2179

Db 3635 GTATTGGTCAGAACAGAAAGGTGATGATCTATCGATTGTGACCCGTGTCATCTGTAGAGG 3694

QY 2180 AAAGGATGATGCAATTGACTAAAGAAAGAAATGGTTCTAGAGCATCTTGTGTTGGG---- 2235

Db 3695 AGAGAATCACTCAGGTTGCGAAAGAAAGAGATGCTTACTCATCTGGTGTACGGCCTG 3754

QY 2236 -----AAACTCAAAACACAAACATTAATCAGGAAGAGTTAGATGACATCATCAGGTATG 2290

Db 3755 GCCTTGGATCAAAAACACTGGCTCCAATGTGCCAACAGGAGCTTGTATGATATCCTCAAGTTTG 3814

QY 2291 GATCAAAAGGAGCTTTTGTAGTGAAGATGATGAAGCAGGAAAGTCTGGAAAAATTCATT 2350

Db 3815 GAACAGAGGAGCTTTTAAAGATGAAGCTACAGAAGGTGGTGAACAAAGAAAGGAGAGG 3874

QY 2351 ATGATGATG-----CGGCTATAGACAAATTGCTTGATCGTGATCTCGTGGAGGCAGAGG 2404

Db 3875 ATATCAGTGTATCCATTACGATGACAGGCCATTGCTCGTCTATTAGACCGGAACGAGG 3934

QY 2405 AAGTCTCAGTGGATGATGAAGAGGAGAAATGGATTCTTAAAGGCTTTCAGGTGGCTAATT 2464

Db 3935 ATGAGACTGAAGAACCAAGAAATTGCAATTGGGAATGAATGAATATCTTAGCTCCTTCAAGG 3994

QY 2465 TTGAATATATAGTGAANAATGAGGCAGCAGCATTAGAGGCACAGAGAGTGGCTGTGAAA 2524

Db 3995 TGGCACAGTATGTGTCCTGCGTGAAGAGGAGATGGGGGATGAAGAGGAATGGTCCGTGAGA 4054

QY 2525 GCAAATCTTCAGCAGGCAATTCTGATAGAGCAAGTATTGGGAAGAGTTGTTAAAAAGATA 2584

Db 4055 TTATTAACAAGAAAGAGTGTAGAT---CCTGATTACTGGGAGAACTTCTCAGGCATC 4111

QY 2585 AATTTGAGCTGCACAGGCTGAGGAGCTTAATGCTCTTGGAAAAAGGAGAGAGTCGCA 2644

Db 4112 ACTATGAGCAGCAGCAGGAAGACCTGGCCGAAATCTAGGCAAGGAAAGCAATACGCA 4171

QY 2645 AGCAGTTGGTATCCATTGAAG-----AAGATGATCTTGCTGCTTTGGAAGATGTGAGCT 2698

Db 4172 AGCAAGTGAACATAATGATGGCAGCCAGGACAGAGATTGGCAGATGATCAGTCTG 4231

QY 2699 CTGATGGAGATGAAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGACAAGGAG 2758

Db 4232 ATAATCAGTCGGATTACTCTGTGGCTTCTGAAGAAGGAGATGAAGATTTTGATGAGAGAT 4291

QY 2759 TTCAGACGGGTCGACGGCCGTACAGAAGAAAGGGTCGCGATAATTGGA-----CCAA 2812

Db 4292 CAGAAGCTGCTCGTCGGCCAAACAGGAAAGGTTTGGGAATGACAAAGACAAAGCCGCTTC 4351

QY 2813 CTCGGTTGATGGAAGGTGAGGGGAGATCTTTCAGAGTACTGGGTTTCAACAGAGTCAAA 2872

Db 4352 CTCCTCTACTGGCTCGTGTGGTGGCAATATAGAGGTGCTTGGATTCAATGCCCGGCAAA 4411

QY 2873 GGGCCATTTTGTACAGACTTTTGATGAGGTATGGAGCTGGCAATTTTGATTGGA----- 2926

Db 4412 GAAAGGCTTTCCTGAATGCCATCATGCGCTATGGGATGCCCCCTCAAGATGCCCTTTACCA 4471

QY 2927 ---AGGAGTTTCTCCTCGCTTAAAGCAGAGACACCTTTGAAGAAATAAATGAATATGGAA 2983

Db 4472 CTCAGTGGCTTGTACAGAGATCTGCGTGGAAATCTGAGAAGGAATTCAGGCTTATGTAT 4531

QY 2984 TACTCTTTGAAGCACATTGCTGAAGAAATAGACGAGAAATTCCTCCACCTTTTCAGATG 3043

Db 4532 CCCTCTTCATCGGCACCTCTGTGAA CCTGGGCAGATGGAGCGGAGACTTTTGCTGATG 4591

QY 3044 GTGTGCCCAAGGAGACTTAGAATAGAAGATGTTCTAGTCAGAAATTCCTTCTGTATAC 3103

Db 4592 GAGTCCCGAGAGAAGGTCTTTCTCGGCAGCATGTGCTAAACAAGAAATTGGCGTCATGTCTC 4651

QY 3104 TAGTTCAGGAGAAAGGT 3119

Db 4652 TGATTCCGCAAAAGGT 4667

RESULT 5

LOCUS BC060721 4343 bp mRNA linear ROD 25-NOV-2003

DEFINITION Mus musculus cDNA clone IMAGE:6406263, partial cds.

ACCESSION BC060721

VERSION BC060721.1 GI:38511411

KEYWORDS

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 4343)

REFERENCE

AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altshul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

TITLE human and mouse cDNA sequences

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

MEDLINE 22388257

REFERENCE

2 (bases 1 to 4343)

AUTHORS Strausberg,R.

TITLE Direct Submission

JOURNAL Submitted (03-NOV-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: Dr. Jim Lin, University of Iowa

cDNA Library Preparation: M. Bento Soares, University of Iowa

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www-shgc.stanford.edu>

Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAC Plate: 125 Row: f Column: 19

This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

FEATURES

source Location/Qualifiers

1. .4343

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6"

/db\_xref="taxon:10090"





Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

2 (bases 1 to 6438)

Strausberg,R.

Direct Submission

Submitted (22-SEP-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: [cgapbs-romail.nih.gov](mailto:cgapbs-romail.nih.gov)

Tissue Procurement: The Cepko Laboratory

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),

Gaithersburg, Maryland;

Web site: <http://www.nisc.nih.gov/>

Contact: [nisc.mgc@nhgri.nih.gov](mailto:nisc.mgc@nhgri.nih.gov)

Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P., Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R., Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAK Plate: 127 Row: m Column: 14

This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, GenomesScan gene prediction, Similarity but not identity to protein.

Location/Qualifiers

1..6438

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="MGC:68269 IMAGE:6489649"

/tissue\_type="Eye, retina, mouse strain C57Bl\6"

/clone\_lib="NIH MGC\_94"

/lab\_host="DH10B"

/note="Vector: pCMV-SPORT6"

1..6438

/gene="Chd4"

/note="synonym: MGC11769"

/db\_xref="LocusID:107932"

/db\_xref="MGI:1344380"

178..5925

/codon\_start=1

/product="chromodomain helicase DNA binding protein 4"

/protein\_id="AAH58578.1"

/db\_xref="GI:35193271"

/db\_xref="LocusID:107932"

/translation="MASGLGSPSPCSAGSEEDMDALLNNSLPPPHPENEEDPDELS

EAETPKLKKKKPKPRDPKIPKSKRQKKELGSDSSGEGEPFVEEEEEVALRSDSEGS

YTPGKKKKKQLGPKKSKSRKEEEEEDEDDDDSKPKSSAQLLEDWGMEDIDHVF

SEEDYRTLTYKAFSQFVRPLIAAKNPKIAVSKMMVLGAKWREFSTNNPFKSGSGAS

VAAAAAAVAVVESMTATEVAPPPPPVEPIRKAKTKBKGPNARRKPKGSPRPVDA

KKPKPKVAPLKI KLGFGSKRKRSSSEDDDDLVESDFDDASINSYSVSDGSTSRSSR

SRKKLRTAKKKKGEEETAVDGYETDHQDYCEVCQGGIILCDTCPRAYHVMVCLDP

DMEKAPEGKWCSPHCEKGIQWEAKEDNSEGEIILEEVGGDEEDDHHMEFCRVCKD

GGELLCCDTCPSSYHIHCLNPPPLPEIPNGEWLCPRCTCPALKGVQKQILIKWKGQPPS

PTPVPRPPDADPNTSPKPLEGRPERQFFVKWQGSYWHCSWVSELQLELHCQVMFRN

YQRKNDMDEPPSGDFGDEEKSRRKNKDPKFAEMEERFYRGIKPEWMMIHRILNHS

VDKKGHVHYLIKWRDLPYDOASWESEDEIODYDFKQSYWNHRELMRGEGRPGKCL

KKVKLRKLERPPETPTVDFTVKYERQPEYLDATGGTILHPYQMEGLNWLRFSWAQGTD

ILADEMGLGKTQVTAVFLYSLYKEGHSKGPFLVASPLSTIINWEREFEMWAPDMYVVT

YVGDKDSRAIIRENEFSFEDNAIRGGKASRMKKEASVKFHVLLTSYELITIDMAILG

SIDWACLIIVDEAHLRKNQSKFFRVNLNGYSLQHKLLLTGTPLQNNLEELFHLNFLT

ERFHNLEGFLEEFADIAKEDIKKLHDMGLGPHMLRRLKADVFKNMPSTELIVRVELS

PMQKKYKYILTRNFEALNARGGNQVSLNVVMDLKKCNHPYLFPPVAAMEAPKMPN

GMVDGSALIRASGKLILLQKMLKNLKEGHRVLIFSQMTKMLDLEDLFLEHEGYKYER

IDGGITGNMQEAIIDRFNAPGAQOFCFLSTRAGGLINLATADTVIYDSDNPHND

IOAFSRAHRIGONKKMIYRFVTRASVEERITQVAKKKMMLTHLVVRPGLSGKTGMS

KQELDDILKFGTEELFDEATDGGDNKEGEDSSVIHYDDKAIERLLDRNQDETEDE

LQGMNEYLSSFKAQYVYREEMEGEEEVEREIIKOBESVDPDYWEKLLRRHHYEQOE

DLARNLGKGRIRKQVNYNDGSQEDRDWQDDSDNSDYSVASEEGEDFDERSEAPR

RPSRKGRLNDKDKPLPILLARVGNIENVLGFNARQKAFINAMRYGMPQDAFTTQW

LVRDLRGKSEKFKAYVSLFMRLHLCPEGADGAETGVPREGLSRQHVLTRIGVMSL

IRKKVQEFEHVNGRWSMPELAEVEENKMSQSPSPKTPPTPTPGDTQNTPAVPVP

AEDGIKEENSLKEESTEGEKEVKSTAEATVECAQAPAPAPATAPATAPEDDDKA

PABPPEGEKEVEKAEKTEEPMETEAKGTTEVEKVEEKSADVLTPIVVEDKEEKK

BEEKDVMLQNGETPKDLSDEKQKNSKQRFMFNIADGGFTELHSLWQNEERAATVTK

KTYEIWRRHDYWLLAGIINHGYARWQDIQNDPRYALINPEFPKGMNRGNFLEIKNKF

LARRFKLLEQALVIEEQLRRAAYLNMSDEPSHPSMALNFRFAEVECLAESHQHLKS

MAGNKPANAVLHKVLKQLEELSDMKADVTRLPATTIARIPPVAVRLQMSERNILSRLA

NRAPEPPPPQOVAQQQ"

1270..1407

/note="PHD; Region: PHD-finger. PHD folds into an

interleaved type of Zn-finger chelating 2 Zn ions in a

similar manner to that of the RING and FYVE domains"

/db\_xref="CDD:pfam00628"

1507..1641

/note="PHD; Region: PHD-finger. PHD folds into an

interleaved type of Zn-finger chelating 2 Zn ions in a

similar manner to that of the RING and FYVE domains"

/db\_xref="CDD:pfam00628"

1774..1887

/note="CHROMO; Region: Chromatin organization modifier

(chromo) domain is a conserved region of around 50 amino

acids found in a variety of chromosomal proteins, which

appear to play a role in the functional organization of

the eukaryotic nucleus. Experimental evidence implicates

the chromo domain in the binding activity of these

proteins to methylated histone tails and maybe RNA. May

occur as single instance, in a tandem arrangement or

followed by a related 'chromo shadow' domain"

/db\_xref="CDD:cd00024"

2020..2184

/note="CHROMO; Region: Chromatin organization modifier

domain"

/db\_xref="CDD:smart00298"

2341..3231

/note="SNF2\_N; Region: SNF2 family N-terminal domain. This

domain is found in proteins involved in a variety of

processes including transcription regulation (e.g., SNF2,

STH1, brahma, MOT1), DNA repair (e.g., ERCC6, RAD16,

RAD5), DNA recombination (e.g., RAD54), and chromatin

unwinding (e.g., ISWI) as well as a variety of other

proteins with little functional information (e.g.,

Iodestar, ETL1)"

/db\_xref="CDD:pfam00176"

3346..3672

/note="HELICC; Region: Helicase superfamily c-terminal

domain"

/db\_xref="CDD:cd00079"

ORIGIN

Query Match		10.2%;	Score 427.4;	DB 10;	Length 6438;
Best Local Similarity		52.9%;	Pred. No. 1.7e-89;		
Matches 1262;		Conservative 0;	Mismatches 1026;	Indels 99;	Gaps 12;
QY	818	TACATCCATACAGCTTGAGGGACTTAATTTTTCGCGTTCTCGTGGTCAAAACAGACGC	877		
Db	2333	TGCATCCCTACCAATGGAAGGCTTAAACTGGCTCCGCTTCTCCTGGGCTCAGGGAACCG	2392		
QY	878	ATGTAATCCTTGCTGATGAAATGGGACTPAGCAAGACAAATTCAAAGCATTCCTCTATCGA	937		
Db	2393	ATACTATCTTGGCTGATGAGATGGCCTTGGGAAAACGTGCGAGACAGCAGTCTTCCCTCT	2452		
QY	938	CTTCACCTTTTGAGGAGAACCTCATTT-----CCGCATTTGGTAATTCCTCTATCGA	991		
Db	2453	ACTCCCTCTATAAGGAGGTCACCTCCAAAGGCCCTTCTTAGTGAGCGCTCCTCTGTCCA	2512		
QY	992	CTCTGCGTAACCTGGGAGAGAGAGTTTGGCCACATGGGCCCCACAGATGAACGTGGTTATGT	1051		
Db	2513	CCATCATCACTGGGAGAGAGAGTTTGAGATGTGGGCTCCTGTATGTATGTGTTAACCT	2572		
QY	1052	ATTTTGGCACTGGCGAAGCTCGAGCAGTTATCAGAGAACATGAGTTTACTTATCGAAAG	1111		
Db	2573	ATGTGGTGACAAGGACAGCCGTGCCATCATCCGGGAAAATGAGTTCTCCTTTGAAGACA	2632		
QY	1112	ATCAAAAAAAGATCAAGAAAAAAGAAATCTGGACAAATAAGTAGCGAAAGCAAGCAAAAA	1171		
Db	2633	ATGCCATTTCGTGGCAAGAGGCCCTC-----GAGGATGAAGAAAGAGCAT	2680		
QY	1172	GAATCAAGTTTGATGTCTCTCCATCGATGATGATGATCAACCTTAGATTTCAGCAGTTC	1231		
Db	2681	CGGTGAATTCATGTTCTGCTGACATCCTATGAGTTGATCACCATTGATATGGCCATCT	2740		
QY	1232	TAAACCAATTAAAGTGGGAGTGATGATGTTGATGAAGGTCTATCGACTGAAAAATAAGG	1291		
Db	2741	TGGGTTCTATTGACTGGGCCCTGCCTCATTTGGATGAAGCCCATCGGCTGAAGAACACC	2800		
QY	1292	ATTCAAAGCTGTTCTCTTCAATTGACACAGTATTCAAAGTAACCAACCGTATTTCTTGACAG	1351		
Db	2801	AGTCTAAGTTCTTCGAGTTTGAATGGTTACTCACTCCAGCACAAAGTGTGCTAACTG	2860		
QY	1352	GAACACCACCTTCAGAACCAACTTGGATGAACCTTTTCATGCTCATGCAATTTCTTGATCGCG	1411		
Db	2861	GAATCCGTTACAGAACAACTAGAGGAACGTGTTTCATCTGCTCAACTTCTCACCCCTG	2920		
QY	1412	GGAAGTTTGGAGTTTGGAGGAGTTCAGGAGGAGTTCCAGGAGGAGTTCAAAGATATTAATCAAGAGGAGC	1471		
Db	2921	AGAGGTTCCACAACTTAGAAGGCTTCTTGGAGGAGTTTGACAGACATTTGCCAAGAGGACC	2980		
QY	1472	AGATCTCAAGTTGTCACAAAAATGTTGGCTCCACATTTGCTCAGAAAGGTAAAAAAGACG	1531		
Db	2981	AGATTAAAAAACTCCACGACATGCTGGGCCCTCATATGTTGGCGGCTCAAAGCTGACG	3040		
QY	1532	TAATGAAAGACATGCCCCCCCCAAAAGGAGCTCATTTTTCGTTGATCTGAGCAGTCTGC	1591		
Db	3041	TCTTCAAGAATATGCCATCCCAAGACAGAGCTGATGTGCGGTGGAGTTGAGTCCCTATGC	3100		
QY	1592	AGAAAGAATATTACAAAGCTATTTTACCCTGAATTAATCAAGTATTGA--CAAAAAAGG	1648		
Db	3101	AGAAGAAGTACTACAGTATATTCTCACCGGGAATTTCGAAGCATTATGCTCGAGGTG	3160		
QY	1649	GAGGTGCTCAAATTTCCCTTAATAACATTTATGATGGAATTACGAAAAGTATGCTGCCATC	1708		
Db	3161	GTGGCAACCAAGGTTTCTCTGCTGAACGTGGTGATGGATCTTAAGAAATGCTGCAACCACC	3220		
QY	1709	CTTA-----TATGCTAGAGGGTGTGAGCCAGTTATTTCACGACGCAAAATGAAGCTT	1759		
Db	3221	CTTATCTCTTCCCTGTGGCAGCAATGGAAGCCCTTAAGATGCCCTAATGGTATGTATGATG	3280		
QY	1760	TCAAAACAACCTTTTGGAGTCTTGTGGAAAAGCTGCAACTTCTAGATAAAATGATGGTCAAAC	1819		
Db	3281	GCAGTGGCCCTAATCAGAGCATCTGGGAAGTTGTGTGCTGTCAGAAGATGTCTTAAGAACCC	3340		

QY	1820	TGAAAGAGCAAGGACACAGAGTCCTTAATATACACACAGTTTCAGCATATGCTGGACTTAC	1879		
Db	3341	TGAAAGAGGGAGGGCATCGTGTGCTCATCTTCTCCAGATGACCAAGATGTTGGACTTGC	3400		
QY	1880	TTGAAGACTACTGTACCCCATAGAAATGGCAGTAGCAGCGAATGATGGAAGGTTGGCG	1939		
Db	3401	TAGAGGATTTTTAGAACATGAAGGTTATAAATATGAACGTTATGATGTTGGAATCACTG	3460		
QY	1940	GAGTGAGCGGCAAAATACGCATAGATCGGTTCAATGCCAAAAATTTAAACAAGTTTGT	1999		
Db	3461	GGACATGCGTCAGGAGGCCATTGATCGCTTCAATGCACCCGGTGTCAACAGTTCTGCT	3520		
QY	2000	TTTTGCTCTCCACAAGAGCTGGTGGCTTAGGAATAAATCTTGCAACGGCTGATACAGTAA	2059		
Db	3521	TCITGCTTCCACTCGAGCTGGGGCCCTGGGATCAATCTGGCCACTGCGGACACAGTTA	3580		
QY	2060	TCATTTATGACAGTGACTGGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC	2119		
Db	3581	TTATATATGACTCTGACTGGAACCCCCATAATGACATCCAGGCTTTAGCAGAGCCCCACC	3640		
QY	2120	GACTTGGCCAAACAAATAAGGTGATGATTTATAGGCTCATAAAACGAGGCACCATTTGAAG	2179		
Db	3641	GTATTTGGGCAAAATAAGAAGGTGATGATCTACCGGTTTGTGCCGTGTCATCAGTGGAGG	3700		
QY	2180	AAAGGATGATGCAATTGACTAAAAAGAAAATGGTTCCTAGAGCATCTTGTGTTGG-----	2234		
Db	3701	AGCGCATCAGCGAGGTGGCAAGAAAGAGATGATGCTGACACATTTAGTGGTACGGCCTG	3760		
QY	2235	----GAAACTCAAAACACAAAACATTAATCAGGAAGAGTTAGATGACATCATCAGGTATG	2290		
Db	3761	GCCTGGGCTCCAAGACAGGCTCAATGTCTAAACAGGAGCTTGATGATATCCTCAAATTTG	3820		
QY	2291	GATCAAGGAGCTTTTGTGCTAGTGAAG-----ATGATGAAGCAG	2329		
Db	3821	GCATCAGGAGCTATTCAAGGATGAAGCCACGGATGGAGGAGAGACAACAAGAGGGAG	3880		
QY	2330	GAAAGTCTGAAAAAATTCATTTATGATGATGCGGCTATAGACAAATGCTTGAATCGTATC	2389		
Db	3881	AAGACAGCAGTGTATCCACTATGACGATAAGGCCATTGAACGACTGCTGGATCGAAACC	3940		
QY	2390	TCGTGGAGGCAGAGGAAGTCTCAGTGGATGATGAAGAGGAGAAATGCGATCTTAAAGGCTT	2449		
Db	3941	AGGATGAGACTGAAGACACAGAAATTG-----CAGGGCATGAATGAATATTGAGCTCAT	3994		
QY	2450	TCAAGGTGCTAATTTTGAATATATAGATGAAAATGAGGCAGCAGCATTAGAGGCACAGA	2509		
Db	3995	TCAAAGTGGCTCAGTATGTGTAACGGGAAGAAAGATGG-----GGGAGGAAGAGG	4045		
QY	2510	GAGTCGCTGCTGAAAAGCAAATCTTCAGCAGGCAATTTCTGATAGCAAGTTATTGGGAAG	2569		
Db	4046	AGGTAGAACGGGAAATCATAAAACAGGAAGAAAGTGTGGATCCTGAC---TACTGGGAGA	4102		
QY	2570	AGTTGTTAAAAAGATAAAATTTGAGCTGCACCCAGGCTGAGGAGCTTAAATGCTCTTGA	2629		
Db	4103	AATTGCTGCGGCACCATTTATGACGACAGCAGCAAGAAAGATCTAGCCCGAAATCTGGC	4162		
QY	2630	GGAAGAGAGTCCCAAGCAGTTGGTATCCATTGAAG--AAGATGATCTTGTGTTTGG	2686		
Db	4163	GAAAAAGAAATCCGTAACACAGGTCAACTACATATGATGGCTCACAGGAGGACCGAGATTGGC	4222		
QY	2687	AAGATGTAGCTCTGTGATGGAGATGAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAG	2746		
Db	4223	AGGACGACCAAGTCCGACCAACCCAGTCCGATTACTCAGTGGCCTCAGAGGAAGGTGATGAAG	4282		
QY	2747	CAGGACAAGGAGTTTCAGACGGGTTCGACGGCCCGTACAGAAGAAAGGG-----TCGCG	2797		
Db	4283	ACTTTGACGAACGGTTCAGAAGCTCCCGCAGGCCCAAGTCGCAAGGGCCTCGGAATGATA	4342		
QY	2798	ATAATTTGGAACCAACTCCGTTGATGGAAGGTGAGGGAGATCTTTTCAAGTACTGGGTT	2857		
Db	4343	AAGATAAGCCATTACCTCCTCTGTTGGCCCGTGTGTTGGTGAATATTTGAAGTACTTGGTT	4402		
QY	2858	TCAACCAGAGTCAAAAGGGCCATTTTGTATACAGACTTTTGTATGATGAGGTATGG-----AG	2908		

Db 4403 TTAATGCTCGTCAGCGAAAGCATTTCTTAATGCAATTATGCGATATGGCATGCCACCTC 4462  
QY 2909 CTGGCAATTTTGAATGGAAGGAGTTTGTTCTCTCGCTTAAAGCAGAGACCTTTGAAGAAA 2968  
Db 4463 AGGATGCTTTTACCACACTCAGTGGCTTTGAGAGATCTTCGAGGCAAGTCAGAGAAAGAA 4522  
QY 2969 TAAATGAATATGGAATACTCTTCTTGAAGCAGCATTTGCTGAAGAAAATAGACGAGAATTCTC 3028  
Db 4523 TTAAGGCTTATGTGTCACTCTTCATCGCACATTTGTGTGAGCCTGGGCGAGATGGGGCTG 4582  
QY 3029 CAACCTTTTCAGATGGTGTGCCAAGGAGGACTTAGAATAGAGATGTTCTAGTCAGAA 3088  
Db 4583 AGACCTTTTCTGATGGCGTCCCACGAGAGGCGCTGTCTCGTCAACATGTTCTCACTAGGA 4642  
QY 3089 TTGCTCTTCTGATACTAGTTTCAGGAGAGAGGTGAAAATTTGTAGAAGAT 3135  
Db 4643 TCGGTGTCATGTCTCTGATTTCGAAAGAGGTTTCAGGAGTTTGAACAT 4689

RESULT 7  
AK128741 6434 bp mRNA linear PRI 09-SEP-2003  
LOCUS Homo sapiens cDNA FLJ16830 fis, clone UTERU3022536, highly similar  
DEFINITION to Chromodomain helicase-DNA-binding protein 4.  
ACCESSION AK128741  
VERSION AK128741.1 GI:34536266  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1  
AUTHORS Kanehori,K., Ishibashi,T., Chiba,Y., Fujimori,K., Hiraoka,S.,  
Tanai,H., Watanabe,S., Ishida,S., Ono,Y., Hotuta,T., Watanabe,M.,  
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,  
Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,  
Kimura,K., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,  
Wagatsuma,M., Takahashi-Fujii,A., Oshima,A., Sugiyama,A.,  
Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K.  
and Isogai,T.  
NEDO human cDNA sequencing project  
Unpublished  
2 (bases 1 to 6434)  
Isogai,T. and Yamamoto,J.  
Direct Submission  
Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7  
Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan  
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
Research Association for Biotechnology (RAB); cDNA library  
construction: Helix Research Institute (HRI) (supported by Japan  
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
HRI, and Biotechnology Center, National Institute of Technology and  
Evaluation; clone selection for full insert sequencing: HRI and  
.RAB; annotation: Reverse Proteomics Research Institute, HRI and  
RAB.

FEATURES  
source Location/Qualifiers  
1..6434  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UTERU3022536"  
/tissue\_type="uterus"  
/clone\_lib="UTERU3"  
/note="Cloning vector: pME18SFL3"

ORIGIN  
Query Match 10.1%; Score 423; DB 9; Length 6434;  
Best local Similarity 52.0%; Pred. No. 1.8e-88;  
Matches 1237; Conservative 0; Mismatches 1075; Indels 69; Gaps 10;

QY 818 TACATCATACCAGCTTGAGGGACTTAATTTTTTGGGTTCTCGTGGTCAAAACAGACGC 877  
Db 2279 TGCACCCCTATCAAATGGAGGCGCTGAATTGGTTGCGTTCTCCTGGGCTCAGGGCACTG 2338  
QY 878 ATGTAATCCTTGCTGATGAAATGGGACTAGGCAAGACAATTCAAAGCATTGCCCTTTTAG 937  
Db 2339 ACGCCATCTTGGCTGATGAGATGGGCTTGGGAAAACTGTACAGACAGCAGTCTTCCTGT 2398  
QY 938 CTTCACTTTTGGAGG-----AACCTCATTTCCGCAATTTGGTAATTCCTCTATCGA 991  
Db 2399 ATTCCCTTTACAAGGAGGTCATTCCAAAGGCCCTTCCTAGTAGCGGCCCTCTTTCTA 2458  
QY 992 CTCTGCGTAACCTGGGAGAGAGATTGCCCACATGGGCCCCACAGATGAACGTGTTATGT 1051  
Db 2459 CCATCATCAAACTGGGAGCGGAGTTTGAATGTGGCTCCAGACATGTATGTCGTAACCT 2518  
QY 1052 ATTTTGGCACTGCGAAGCTCGAGAGTTATCAGAGAACATGATGTTTACTTATCGAAAG 1111  
Db 2519 ATGTGGGTGACAAAGGACAGCCGTGCCATCATCCGAGAGAAATGATGTTCTCTTGAAGACA 2578  
QY 1112 ATCAAAAAAAGATCAAGAAAAAGAAATCTGGACAAATAAGTAGGAAAGCAAGCAAAAAA 1171  
Db 2579 ATGCCATTCGTGGTGGCAAGAGGCCCTC-----CCGCATGAAGAAAGAGGCAT 2626  
QY 1172 GAATCAACTTTGATGTCTCTCCATCAGATCGTATGAGATGATCAACCTAGATTTCAGCAGTTC 1231  
Db 2627 CTGTGAATTCATGTGCTGCTGACATCCTATGAATTGATCACCATTGACATGGCTATTT 2686  
QY 1232 TAAACCAATTAAGTGGGAGTGCATGATTGTTGATGAAGGTTCATCGACTGAAAAATAAGG 1291  
Db 2687 TGGGCTTATTGATTGGGCGCTGCCCTCATCGTGGATGAAGCCCATCGGCTGAAGACAATC 2746  
QY 1292 ATTCAAAGCTGTTCTCTTCATTGACACAGTATTCAAAGTAACCAACCGTATTCTTCTGACAG 1351  
Db 2747 AGTCTAAGTTCTTCGGGTATTGAATGGTTACTCACTCCAGCACAAGCTGTTGCTGACTG 2806  
QY 1352 GAACACCACTTCAGAACAACTTGGATGAACCTTTTCATGCTCATGCAATTTCTTGATCGG 1411  
Db 2807 GGACACCATTACAAAAACAATCTGGAAGAGTTGTTTCATCTGCTCAAACTTCTCACCCCG 2866  
QY 1412 GGAAGTTTGGAAGTTTGAGGAGTTCCAGGAGGAGTTCAAAGATATTAAATCAAGAGGAGC 1471  
Db 2867 AGAGGTTCCACAATTTGGAAGGTTTTTTGGAGGAGTTTGCTGACATTGCCAAGGAGGACC 2926  
QY 1472 AGATCTCAAGGTTGCACAAAAATGTTGGCTCCACATTTGCTCAGAAGGGTAAAAAAGACG 1531  
Db 2927 AGATAAAAAAACTGCATGACATGCTGGGCGGCACATGTTGCGGCGGCTCAAAGCCGATG 2986  
QY 1532 TAATGAAAGACATGCCCCCCCCAAAAAGGAGCTCATTTTGGCTGTTGATCTGAGCAGTCTGC 1591  
Db 2987 TGTTCAAGAACATGCCCTCCAGGACAGAACTAATTGTGCGTGTGGAGCTGAGCCCTATGC 3046  
QY 1592 AGAAGAAATATTACAAAGCTATTTTACCCGTAATTATCAAGTATTGA---CAAAAAAGG 1648  
Db 3047 AGAAGAAATACTACAAGTACATCCTCACTCGAAATTTTGAAGCACTCAATGCCCGAGGTG 3106  
QY 1649 GAGGTGCTCAAAATTTCCCTTAATAACATATTATGATGGAATTACGAAAAAGTATGCTGCCATC 1708  
Db 3107 GTGGCAACCAGGTGTCTCTGCTGAATGTGGTGATGGATCTTTAAGAAGTGCTGCAACCATC 3166  
QY 1709 CTTA-----TATGCTAGAGGGTGTGAGCCAGTTATTACGACGCAAAATGAAGCTT 1759  
Db 3167 CATACCTCTTCCCTGTGGCTGCAATGGAAGCTCCTAAGATGCCTAATGGCATGTATGATG 3226  
QY 1760 TCAAAACAACCTTTTGGAGTCTTGTGGAAAGCTGCAACTTCTAGATAAAATGATGGTCAAAC 1819  
Db 3227 GCAGTGCCCTAATCAGAGCATCTGGGAAATATTGCTGCTGCAGAAAAATGCTCAAGAACC 3286  
QY 1820 TGAAAGAGCAAGGACACAGAGTCCTTAATATACACACAGTTTCAGCATATGCTGGACTTAC 1879  
Db 3287 TTAAGGAGGGTGGGCATCGTGTACTCATCTTTTCCAGATGACCAAGATGCTAGACCTGC 3346  
QY 1880 TTGAAGACTACTGTACCCTACCAATAAGAAATGGCAGTACGAGCGGAATTGATGGAAGGTTGGCG 1939



Db 3347 TAGAGGATTTCTTGGAACATGAAGGTTATAAATAACGAACGCATCGATGGTGAATCACTG 3406  
Qy 1940 GAGCTGAGCGGCAAAATACGCATAGATCGGTTCAATGCCAAATAATCTTAACAAAGTTTTGTT 1999  
Db 3407 GGAACATGCGGCAAGAGGCCATTGACCGCTTCAATGACCGGTGCTCAGCAGTTCTGCT 3466  
Qy 2000 TTTTGTCTCCACAAGAGCTGGTGCTTAGGAATAAAATCTTGAACGGCTGATACAGTAA 2059  
Db 3467 TCITGCTTCCACTCGAGCTGGGGCCTTGGAAATCAATCTGGCACTGCTGACACAGTTA 3526  
Qy 2060 TCATTTATGACAGTGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC 2119  
Db 3527 TTATCTATGACTCTGACTGGAACCCCATAAATGACATTCAGGCTTTAGCAGAGCTCAC 3586  
Qy 2120 GACTTGGCCAAACAAATAAGGTGATGATTTATAGGCTCATAAACCGAGGCCACCATTTGAAG 2179  
Db 3587 GGATTGGGCAAAATAAAAAGGTAATGATCTACCGGTTTGTGACCGTGCGTCAGTGGAGG 3646  
Qy 2180 AAAGGATGATGCAATTGACTAAAAAGAAAATGGTTCTAGAGCATCTTGTGTTGG----- 2234  
Db 3647 AGCGCATCAGCAGGTGGCAAGAAAGAAAATGATGCTGACGCATCTAGTGTGCGGCCTG 3706  
Qy 2235 ----GAAACTCAAAACACAAAACATTAATCAGGAAGAGTTAGATGACATCATCAGGTATG 2290  
Db 3707 GGCTGGGCTCCAAGACTGGATCTATGTCCAACAGGAGCTTGATGATATCCTCAAATTTG 3766  
Qy 2291 GATCAAAAGGAGCTTTTGTCTAGTGAAGATGATGAAGCAGGAAGTCTGGAATAATTCAT 2350  
Db 3767 GCACTGAGAACTATTCAAGGATGAAGCCACTGATGGAGGAGGAGACAACAAGAGGGAG 3826  
Qy 2351 ATGATGATGCGGCTATAGACAAATTTGCTGA-----TCGTGATCTCGTGGAGGCAGAGG 2404  
Db 3827 AAGATAGCACTGTTATCCACTACGATGATAAGGCCATTGAACGGCTGCTAGACCGTAACC 3886  
Qy 2405 AAGTCTCAGTGGATGATGAAGAGGAGAAATGGATTTCTAAAGGCTTTCAGGTGGCTAAT 2464  
Db 3887 AGGATGAGACTGAAGACACAGAATTGCAGGGCATGAATGAATATTGAGCTCATTTCAAAG 3946  
Qy 2465 TTGAATATATAGATGAAAATGAGGCAGCAGCATTAGAGGCACAGAGATCGTCTGTA 2524  
Db 3947 TGGCCCACTATGTGGTACGGGAAGAGAAATGGGGAGGAAGAGGAGGTAGAACGGGAAA 4006  
Qy 2525 GCAAATCTTCAGCAGGCAATTCTGATAGAGCAAGTTATTGGGAAGAGTTGTTAAAGATA 2584  
Db 4007 TCATTAACAGGAAGAAAGTGTGGATCTCTGA---CTACTGGGAGAAATTGCTGCGGCACC 4063  
Qy 2585 AATTTGAGCTGCACCGCTGAGGAGCTTAATGCTCTTGGAAGAAAGAGAGAAAGTCGCA 2644  
Db 4064 ATTATGAGCAGCAGCAAGAAGATCTAGCCCCGAAATCTGGGCAAGGAAAGAAATCCGTA 4123  
Qy 2645 AGCAGTTGGTATCCATTGAAG--AAGATGATCTTGTCTGGTTTGAAGATGTGAGCTCTG 2701  
Db 4124 AACAGGTCAACTACAAATGATGGCTCCAGGAGGACCGAGATTGGCAGGACGACCACTCCG 4183  
Qy 2702 ATGGAGATGAAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGACAAGGAGTTC 2761  
Db 4184 ACAACCACTCCGATTACTCAGTGGCTTCAGAGGAAGGTGATGAAGACTTTGATGAACGTT 4243  
Qy 2762 AGACGGTTCGACGCCGTACAGAAGAAAGGTCGCGATAATTTGGAAACCAA----- 2812  
Db 4244 CAGAAGCTCCCCGTAGGCCCAGTCGTGAAGSCCTGCGGAATGATAAAGATAAGCCATTGC 4303  
Qy 2813 CTCGGTTGATGGAAGGTGAGGGGAGATCTTTCAGAGTACTGGGTTTCAACACAGAGTCAAA 2872  
Db 4304 CTCCTCTGTTGGCCCGTGTGGTGGGAATATTGAAGTACTTGGTTTAAATGCTCGTCAGC 4363  
Qy 2873 GGGCCATTTTGTACAGACTTTTGATGAGGTATGG-----AGCTGGCAATTTTGATT 2923  
Db 4364 GAAAAGCCTTTCTTAATGCAATTATGCGGTATGGTATGCCACCTCAGGATGCTTTTACTA 4423  
Qy 2924 GGAAGGAGTTTGTCTCGCTTAAAGCAGAGACCTTTGAAGAAAATAAATGAATATGGAA 2983

Db 4424 CCCAGTGGCTTGTAAAGAGACCTGCGAGGCAAAATCAGAGAAAAGAGTTCAAGGCATATGTCT 4483  
Qy 2984 TACTCTTCTTTGAAGCACATTTGCTGAAGAAAATAGACGAGAATTTCTCCAACCTTTTTCAGATG 3043  
Db 4484 CTCCTTTTCATCGGCATTTATGTGAGCCGGGCGCAGATGGGCTGAGACCTTTTGTCTGATG 4543  
Qy 3044 GTGTGCCCCAAGGAAGGACTTAGAATAGAAATAGAGATGTTCTAGTCAGAAATTGCTCTCTGATAC 3103  
Db 4544 GTGTCCCCCGAGAAGGCCTGTCTCGCCAGCATGTCTTACTAGAAATGGTGTATGTCTT 4603  
Qy 3104 TAGTTCAGGAGAAGGTGAAATTTGTAGAAGATCATCCAGGG 3144  
Db 4604 TGATTGCAAGAAGGTTCAGGAGTTTGAACATGTTTAATGGG 4644

RESULT 8  
A56817  
LOCUS A56817 6327 bp DNA linear PAT 03-MAR-1998  
DEFINITION Sequence 1 from Patent WO9628547.  
ACCESSION A56817  
VERSION A56817.1 GI:3712830  
KEYWORDS .  
SOURCE unidentified  
ORGANISM unidentified  
REFERENCE 1  
AUTHORS Seelig,H.P. and Renz,M.  
TITLE DERMATOMYOSITIS-SPECIFIC AUTO-ANTIGEN  
JOURNAL Patent: WO 9628547-A 1 19-SEP-1996;  
PRIVATE INST FUER IMMUNOLOGIE (DE)  
COMMENT Other publication DE 19509279 960515.  
FEATURES  
Location/Qualifiers  
1..6327  
/organism="unidentified"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32644"  
<1..5738  
/note="unnamed protein product"  
/codon\_start=3  
/protein\_id="CAA03424.1"  
/db\_xref="GI:3712831"  
/db\_xref="REMTREMBL:CAA03424"  
/translation="ASGLSPSPCSAGSEEDMDALLNNSLPPHPHNEEDPEEDLSE  
TETPKLKKKKPKPRDPKIPKSKRQKKERMILCRQLGDSGGEPEFVEEEVVALRS  
DSEGDYTPGKKKKKLGPKKEKSKSRKEEEDDDDDSKPKSSAQLLEDWME  
DIDHVFSEEDYRTLNTYKAFSQFVRPLIAAKNPKIAVSKMMVVLGAKWREFSTNPPK  
GSSGASVAAAAAVALVESMVTATEVAPPPVEVPVIRKAKTKEGKGNRRKPKGS  
PRVPDAKKPKKVPAPLKILGGFGSKRRSSDDDLDESDFDDASINSYSVSDGS  
TSRSSRSRKKLRTTKKKKKGEEEVAVDGYETHDQYCEVCQQGGEIILCDTCPRAYH  
MVCLDPDMEKAPEGKWSCPHCEKEGIQWEAKENSGEEILEEVGGDLEEEDHHMEF  
CRVCKDGGELCCDTPSSYHIHCLNPPPEINPGEWELCPRCTCPALKGKVQKILIWK  
WGQPPSPTVPVRPPDADPNTSPKPLEGRPERQFVKWQMSYWHCSWVSELQLELHC  
QVMFRNYQRKNDMDPPSGDFGGDEKSKRKNKDPKFAEMEERFYRYGIKPEWMMIH  
RILNHSVDKKGHVYLIKWRDLPYDQASWESEVEIQDYDLFKQSYWNHRELMRGEEG  
RPGKKLKKVLRKLERPPETPTVDPTVKYEROPEYIDATGCTLHPYQMEGLNWLRFWS  
AQGTDITILADEMGLKTVQTAFLYSLYKEGHSKGFVLSAPLSTIINWEREFEMWAP  
DMYVVTYVGDKDSRAIIRENEFSFEDNAIRGGKASRMKKEASVKFHVLLTSYELITI  
DMAILGSIIDWACLIIVDEAHLKNNQSKFFRVNLNGYSLQHKLLLTGTPLQNNLEELFHL  
LNFLTPERFHNLEGLEEFADIAKEDQIKKLHMLGPHMLRRLKADVFNKMPSTELI  
VRVELSPMQKKYKIILTRNFENALNARGGNQVSLNVNMDLKKCCNHPYLPFVAAME  
APKPNMGYDGSALIRASGKLLLLQKMLKNLKEGHRVLIIFSQMTKMLDILLEDLEHE  
GKYERIDGGITGNMQEAIIDRFNAPGAQQFCFLSTRAGGLGINLATADTVTIYDS  
WNPNDIQAFSRAHRIQGNKKVMIYRFVTRASVEERITQVAKKKMLTLHVVRLGSG  
KTGSMKQELDDILKFGTEELFKDEATDGGDNKGBEDSSVIHYDDKAIERLLDRNQD  
ETEDTELQGMNEYLSFKVAQYVVRREEEMGEEVEEREIIKQESVDPDYWEKLLRHH  
YEQQOEDLARNLKGKRIKQVNYNDGSEDREDWQDDQSDNQSDSVASEEGEDDFDE  
RSEAPRRPSRKGLRNDKDKPLPLLARVGNIEVLGNARQKAFNAIMRYGNPQQD  
AFTTQWLVRDLRGKSEKEFKAYVSLFMRHLCEPADGAETFADGVPREGLSRQHVLT  
IGWMSLIRKKVQFEFHVNGRWSMPELAEVEENKMSQPSPKPTPTSPSTPGTQPT  
PAPVPPAEDGIKIENSLKEEESI EGEKEVKSTAPETAI ECTQAPAPASEDEKVVVEP  
PEGEKEVKAEVKERTPEPMEETPKGAADVKEVBKSAIDLTPIVVEDKEEKEEEK  
KEVMLQNGETPKDLNDEKQKNIKQRFMFNIADGGFTELHSLWQNEERAATVTKTYE  
IWHRRHDYWLLAGIINHGYARWQDIQNDPRYAILNPFKGMNRGNFLEIKNKLARR

FKLLEQALVIEEQLRRAAYLNMSEDPSHPMSMALNTRFAEVEBCLAESHQHLKSKESMAGN  
KPANAVLHKVLKQLEELLSDMKADVTRLPATIARIPPPAVRLQMSERNILSRLANRAP  
EPTPQOVAQQQ"

ORIGIN

Query Match 10.1%; Score 421.4; DB 6; Length 6327;

Best Local Similarity 51.9%; Pred. No. 4.4e-88;

Matches 1236; Conservative 0; Mismatches 1076; Indels 69; Gaps 10;

Qy	818	TACATCCATACCAAGCTTGAGGGACTTAATTTTTCGCGTTCCTCGTGGTCAAAAACAGACGC	877
Db	2176	TGCACCCCTATCAAAATGGAGGGCCTGAATTGGTTGCGCTTCTCCTGGGCTCAGGCACTG	2235
Qy	878	ATGTAATCCTTGCTGATGAAATGGGACTAGGCAAGACAAATTCAAAGCATTTGCCCTTTTAG	937
Db	2236	ACACCATCTTGGCTGATGATGGCCCTTGGGAAAACTGTACAGACAGCAGTCTTCCTGT	2295
Qy	938	CTTCACTTTTGGAGG-----AACCTCATTCGCGATTGTTGGTAATTGCTCCTCTATCGA	991
Db	2296	ATTCCCTTTACAGGAGGGTCAATCCAAAGGCCCTTCCTAGTGAGCGGCCCTCTTTCTA	2355
Qy	992	CTCTGCGTAACCTGGAGAGAGAGTTTGGCCACATGGGCCCCACAGATGAACGTGGTTATGT	1051
Db	2356	CCATCATCAACTGGAGCGGGAGTTTGAATGTGGGCTCCAGACATGTATGTTCGTAACCT	2415
Qy	1052	ATTTTGGCACTGCGCAAGCTCGAGCAGTTATCAGAGAACATGAGTTTTTACTTATCGAAAG	1111
Db	2416	ATGTGGGTGACAAAGGACAGCCGTGCCATCATCCGAGAGAATGAGTTCCTCTTGAAGACA	2475
Qy	1112	ATCAAAAAAGATCAAGAAAAAAGNAATCTGGCAAAATAAGTAGCGAAAGCAAGCAAAAA	1171
Db	2476	ATGCCATTTCGTGGTGCAAGGAAGGCCCTC-----CCGCATGAAGAAAGAGGCAT	2523
Qy	1172	GAATCAAGTTTGATGTCCTCCTCACATCGTATGAGATGATCAACCTAGATTTCAGCAGTTC	1231
Db	2524	CTGTGAAATTCCATGTGCTGACATCCTTATGAATTGATCACCATTGACATGGCTATTT	2583
Qy	1232	TAAACCAATTAAGTGGGAGTGATGATTGTTGATGAAGGTCATCGACTGAAAAATAAGG	1291
Db	2584	TGGGCTCTATTGATTGGGCCTGCCTCATCGTGGTAGAGCCCATCGGCTGAAGAACAAATC	2643
Qy	1292	ATTCAAAGCTGTTCTCTTCAATTGACACAGATTAACAAGTAACCAACCGTATTCTTCTGACAG	1351
Db	2644	AGTCTAAGTTCTTCGGGTATTGAATGGTTACTCACTCCAGCACAAAGCTGTGCTGACTG	2703
Qy	1352	GAACACCACCTTCAGAACAACTTGGATGAACTTTTCATGCTCATGCAATTTCTTGATCGCG	1411
Db	2704	GGACACCATTACAAAACAATCTGGAAGAGTTGTTTCATCTGCTCAACTTTCTCACCCCGG	2763
Qy	1412	GGAAGTTTGAAGTTTGGAGGAGTTCCAGGAGGAGTTCCAGAGATATTAATCAAGAGGAGC	1471
Db	2764	AGAGGTTCCACAATTTGGAAGGTTTTTTGGAGGAGTTTGTGACATTGCCAAGGAGGACC	2823
Qy	1472	AGATCTCAAGTTTGCAAAAAATGTTGGCTCCACATTGCTCAGAAGGGTAAAAAAGACG	1531
Db	2824	AGATAAAAAAACTGCATGACATGCTGGGGCCGACATGTTGCGCGGCTCAAAGCCGATG	2883
Qy	1532	TAATGAAAGACATGCCCCCAAAAAGGAGCTCATTTTGGGTGTTGATCTGAGCAGTCTGC	1591
Db	2884	TGTTCAAGAACATGCCCTCCAAGACAGAACTAATTGTGCGTGTGGAGCTGAGCCCTATGC	2943
Qy	1592	AGAAAGAATATTACAAGCTATTTTTTACCCGTAATTATCAAGTATTGA---CAAAAAAGG	1648
Db	2944	AGAAGAAATACTACAAGTACATCCTCACTCGAAATTTTGAAGCACTCAATGCCCCGAGGTG	3003
Qy	1649	GAGGTGCTCAAAATTTCCCTTAATAACATTAATGATGGAATTAACGAAAAAGTATGTCGCATC	1708
Db	3004	GTGGCAACCAGGTGCTCTGCTGAATGTGGTATGATGATCTTAAGAAGTGTCTGCAACCATC	3063
Qy	1709	CTTA-----TATGCTAGAGGGTGTGTAGCCAGTTATTTCAGCACGCAATGAAGCTT	1759
Db	3064	CATACCTCTTCCCTGTGGCTGCAATGGAAGCTCTTAAGATGCCTAATGGCATGTATGATG	3123

Qy	1760	TCAACAACACTTTTGGAGTCTTGTGGAAAAGCTGCAACTTCTAGATAAAATGATGGTCAAAAC	1819
Db	3124	GCAGTGCCCTAATCAGAGCATCTGGGAAATATTATTGCTGTCAGAAAATGCTCAAGAACC	3183
Qy	1820	TGAAGAGCAAGGACACAGAGTCTCTAATATATACACACAGTTTCAGCATATGCTGGACTTAC	1879
Db	3184	TTAAGGAGGTGGGCATCGTGTACTCATCTTTTCCAGATGACCAAGATGCTAGACCTGC	3243
Qy	1880	TTGAAGACTACTGTACCCATAAGAAATGGCAGTACGAGCGAATTGATGGAAAAGTTGGCG	1939
Db	3244	TAGAGGATTCTTGGAAACATGAAGGTTATAAATACGAACGCATCGATGGTGAATCACTG	3303
Qy	1940	GAGCTGAGCGGCAAAATACGCATAGATCGGTTCAATGCCAAAAATTTTAAACAAGTTTGT	1999
Db	3304	GGAACATGCGGCAAGAGGCCATTGACCGCTTCAATGCACCCGGGTGCTCAGCAGTTCTGCT	3363
Qy	2000	TTTTGCTCTCCAAAGAGCTGGTGGCTTAGGAATAAATCTTGCAACGGCTGATACAGTAA	2059
Db	3364	TCCTGCTTCCACTCGAGCTGGGGCCTTGAATCAATCTGGCCACTGCTGACACAGTTA	3423
Qy	2060	TCATTTATGACAGTGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGTCACT	2119
Db	3424	TTATCTATGACTCTGACTGGAACCCCCATAATGACATTCAGGCCCTTAGCAGAGCTCACC	3483
Qy	2120	GACTTGGCCAAAACAAATAAGGTGATGATTTATAGGCTCATAAACGAGGCACCATTTGAAG	2179
Db	3484	GGATTGGSCAAAATAAAAAGGTAATGATCTACCGTTTGTGACCCGTCGTGAGTGAGG	3543
Qy	2180	AAAGGATGATGCAATTGACTAAAAAGAAAAATGTTCTAGAGCATCTTGTGTTGG-----	2234
Db	3544	AGCGCATCAGCAGGTGGCAAAGAGAAAAATGATGCTGACGCATCTAGTGTGCGGCCCTG	3603
Qy	2235	----GAAACTCAAAACACAAACATTAATCAGGAAGAGTTAGATGACATCATCAGGTATG	2290
Db	3604	GGCTGGGCTCCAAGACTGGATCTATGTCCAAACACAGGAGCTTGTATGATATCTCAAATTTG	3663
Qy	2291	GATCAAAGGAGCTTTTGTAGTGAAGATGATGAAGCAGGAAAGTCTGGAATAAAATTCATT	2350
Db	3664	GCACTGAGGAACATTTCAAGGATGAAGCCACTGATGGAGGAGGAGACAAACAAAGAGGAG	3723
Qy	2351	ATGATGATGCGGCTATAGACAAAATTGCTTGA-----TCGTGATCTCGTGAGGCAGAGG	2404
Db	3724	AAGATAGCAGTGTATTCCACTACGATGATAAGGCCATTGAAACGGCTGCTAGACCGTAACC	3783
Qy	2405	AAGTCTCAGTGGATGATGAAGAGGAGAAATGGATTCTTAAAGGCTTTCAAGGTGGCTAATT	2464
Db	3784	AGGATGAGACTGAAGACACAGAATTCAGGGCATGAATGAATATTTGAGCTCATTCAAAG	3843
Qy	2465	TTGAATATATAGATGAAATGAGGCAGCAGCATTTAGAGGCACAGAGAGTCGCTGAAA	2524
Db	3844	TGGCCCACTATGTGGTACGGGAAGAGAAATGGGGAGGAAGAGAGGTAGAACGGGAAA	3903
Qy	2525	GCAAACTTTCAGCAGGCAATTTCTGATAGAGCAAGTTATTGGGAAGAGTTGTTAAAGATA	2584
Db	3904	TCATTAAACAGGAAGAAAGTGTGGATCCTGA---CTACTGGGAGAAATTTGCTGCGGCACC	3960
Qy	2585	AATTGAGCTGCACCAAGGCTGAGGAGCTTAATGCTCTTTGGAAAAAGGAAGAGAGTCGCA	2644
Db	3961	ATTATGAGCAGCAGCAAGAAGATCTAGCCGAAATCTGGGCAAAAGGAAAAAGAAATCCGTA	4020
Qy	2645	AGCAGTTGGTATCCATTGAAG---AAGATGATCTTGTGTTTGGAAAGATGTGAGCTCTG	2701
Db	4021	AACAGGTCAACTACAATGATGGCTCCCAGGAGGACCGAGATTGGCAGGACGACCAAGTCCG	4080
Qy	2702	ATGGAGATGAAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGACAAAGGAGTTC	2761
Db	4081	ACAACCAGTCCGATTACTCAGTGGCTTCAGAGGAAGGTGATGAAGACTTTTGATGAACGTT	4140
Qy	2762	AGACGGGTGACGGCCGTACAGAAGAAAGGTCGCGGATAATTTGGAAACCA-----	2812
Db	4141	CAGAAGCTCCCGTAGGCCCAAGTCGTAAGGCCCTGCGGAATGATAAAGATAAGCCATTGC	4200
Qy	2813	CTCCGTTGATGGAAGGTGAGGGGAGATCTTTCAGAGTACTGGGTTTCAACCAGAGTCAAA	2872



Db	4201	CTCCTCTGTTGGCCCGTGTGGTGGGAATATTGAAGTACTTGGTTTAAATGCTCGTCAGC	4260
Qy	2873	GGGCCAATTTTGTACAGACTTTTGATGAGGTATGG-----AGCTGGCAATTTTGATT	2923
Db	4261	GAAAAGCCTTCTTAATGCAATTATGCGATATGGTATGCCACCTCAGGATGCTTTTACTA	4320
Qy	2924	GGAAGGAGTTTGTCTCGCTTAAAGCAGAAGACCTTTTGAAGAAATAAATGAATATGGAA	2983
Db	4321	CCCAGTGGCTGTGAAGAGACCTGCGAGGCCAAATCAGAGAAAGAGTTCAAGGCATATGTCT	4380
Qy	2984	TACTCTTTGAAGCACATTTGCTGAAGAAATAGACGAGAATTTCTCCAACCTTTTCAGATG	3043
Db	4381	CTCTTTTCATGCGGCATTTATGTAGCGCGGGGCAGATGGGGCTGAGACCTTTGCTGATG	4440
Qy	3044	GTGTGCCCAAGGAGGACTTAGAATAGAAGATGTTCTAGTCAGAAATTGCTTCTGATAC	3103
Db	4441	GTGTCCCGGAGAAAGGCCTGTCTGCCAGCATGTCCTTACTAGAAATGGTGTATGTCTT	4500
Qy	3104	TAGTTCAGGAGAAGGTGAAATTTGTAGAAGATCATCCAGGG	3144
Db	4501	TGATTCGCAAGAAGGTTTCAGGAGTTTGAACATGTTAATGGG	4541
RESULT 9			
AR224019			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
1. .6328			
/organism="unknown"			
/mol_type="genomic DNA"			
ORIGIN			
Query Match			
Best Local Similarity			
Matches 1236; Conservative			
0; Mismatches 1076; Indels 69; Gaps 10;			
Qy	818	TACATCCATACAGCTTGAGGGACTTAATTTTTCGGGTTCTCGGTCAAAACAGACGC	877
Db	2177	TGCACCCCTATCAAAATGGAGGGCCTGAATTGGTTGCGCTTCTCCTGGCTCAGGGCAGT	2236
Qy	878	ATGTAATCCTTGCTGATGAAATGGGACTAGGCAAGACAAATTCAAAGCATTCCTTTTAG	937
Db	2237	ACACCATCTTGGCTGATGAGATGGGCCCTTGGGAAACTGTACAGACAGCAGTCTTCCTGT	2296
Qy	938	CTTCACCTTTTGGAGGAG-----AACCTCATTCGCGATTTGGTAATTGCTCCTATCGA	991
Db	2297	ATTCCTTTTACAAGGAGGGTCAATTCCAAAGCCCTTCTCCTAGTAGAGCGCCCTCTTTCTA	2356
Qy	992	CTCTGCGTAACCTGGGAGAGAGAGTTTGCACATGGGCCCCACAGATGAACGTGGTTATGT	1051
Db	2357	CCATCATCAACTGGGAGCGGAGTTTGAATGTGGGCTCCAGACATGATGTGTAACCT	2416
Qy	1052	ATTTTGGCACTGGCCAAGCTCGAGCAGTTATCAGAGAACATGAGTTTACTTATCGAAAG	1111
Db	2417	ATGTGGTGACAAAGACAGACCGCTGCCATCATCCGAGAGAATGAGTTCTCCTTTGAAGACA	2476
Qy	1112	ATCAAAAAAAGATCAAGAAAAAGAAATCTGGACAAATAAGTAGCGAAAGCAAGCAAAAA	1171
Db	2477	ATGCCATTCTGTGGTGGCAAGAGGCCTC-----CCGCATGAAGAAAGAGGCAT	2524
Qy	1172	GAATCAAGTTTGATGTCCTCCTCACATCGTATGAGATGATCAACCTAGATTGAGCAGTTC	1231

Db	2525	CTGTGAAATTCATGTGCTGCTGACATCCTATGAATTGATCACCATTGACATGGCTATTT	2584
Qy	1232	TAAAAACCAATTAAGTGGGAGTGCATGATTGTTGATGAAGGTCACTCGACTGAAAAATAAGG	1291
Db	2585	TGGGCTCTATTGATTGGGCTGCTCATCGTGGATGAAGCCCATCGGCTGAAGAACAAATC	2644
Qy	1292	ATTCAAAAGCTGTTCTCTTCAATTGACACAGTATTCAAAGTAACACACCGTATTCTTCTGACAG	1351
Db	2645	AGTCTAAGTTCTTCCGGGTATTGAATGGTTACTCACTCCAGCACAAAGCTGTTGCTGACTG	2704
Qy	1352	GAACACCACTTCAGAACAACTTGGATGAACCTTTTCATGCTCATGCAATTTCTTGTATGCGG	1411
Db	2705	GGACACCATTAACAAAACAATCTGGAAGAGTTGTTTCATCTGCTCAACTTCTCACCCCGG	2764
Qy	1412	GGAACTTTGGAAGTTTGGAGGAGTTCCAGGAGGAGTTCAAAGATATTAAATCAAGAGGAGC	1471
Db	2765	AGAGGTTCCACAATTTGGAAGGTTTTTTGGAGGAGTTTGTGCTGACATTGCCAAGGAGGACC	2824
Qy	1472	AGATCTCAAGGTTGCACAAAATGTTGGCTCCACATTTTGTCTCAGAAAGGTAAAAAAGACG	1531
Db	2825	AGATAAAAAAACTGCATGACATGCTGGGGCCGCACATGTTGGCGGGCTCAAAGCCGATG	2884
Qy	1532	TAATCAAAAGACATGCCCCCCTCAAAAGAGGAGCTCATTTTGGTGTGATCTGAGCAGTCTGC	1591
Db	2885	TGTTCAAGAACATGCCCCCTCCAAGACAGAACTAATTTGCGTGTGGAGCTGAGCCCTATGC	2944
Qy	1592	AGAAAGAAATATTACAAAGCTATTTTACCCCGTAATTTATCAAGTATTGA---CAAAAAAGG	1648
Db	2945	AGAAAGAAATACTACAAGTACATCCTCACTCGAAATTTTGAAGCACTCAATGCCGAGGTG	3004
Qy	1649	GAGGTGCTCAAATTTCCCTTAATAACATTTATGATGGAATTACGAAAAAGTATGCTGCCATC	1708
Db	3005	GTGGCAACCAGGTGCTCTCTGCTGAATGTGGTGATGGATCTTAAGAAGTGTGCAACCATC	3064
Qy	1709	CTTA-----TATGCTAGAGGTGTTGAGCCAGTTTATTCACGACGCAATGAAGCTT	1759
Db	3065	CATACCTCTTCCCTGTGGCTGCAATGGAAGCTCCTAAGATGCCTAATGGCATGTATGATG	3124
Qy	1760	TCAAAACAACCTTTTGGAGTCTTGTGGAAAGCTGCAACTTCTAGATAAAATGATGGTCAAAAC	1819
Db	3125	GCAGTGCCCTAATCAGAGCATCTGGGAAATTTATTGCTGTGCGAGAAATGCTCAAGAACCC	3184
Qy	1820	TGAAAGAGCAAGGACACAGAGTCTTAATAATATACACACAGTTTTCAGCATATGCTGGACTTAC	1879
Db	3185	TTAAGGAGGTGGGCATCGTGTACTCATCTTTTCCCAGATGACCAAGATGCTAGACCTGC	3244
Qy	1880	TTGAAGACTACTGTACCCATAAGAAATGGCAGTACGAGCGCAATTTGATGGAAGGTTGGCG	1939
Db	3245	TAGAGGATTTCTTGGAAACATGAAGTTATAAATACGAACGCATCGATGGTGAATCACTG	3304
Qy	1940	GAGCTGAGCGGCAAAATACGCATAGATCGGTTCAATGCCAAAAATTTCTAACAAAGTTTGT	1999
Db	3305	GGACATGCGGCAAGAGGCCATTGACCGCTTCAATGCACCGGGTGTCTAGCAGATTTCTGCT	3364
Qy	2000	TTTTGCTCTCCACAAGAGCTGGTGGCTTAGGAATAAATCTTGCAACGGCTGATACAGTAA	2059
Db	3365	TCCTGCTTCCACTCGAGCTGGGGCTTGGAAATCAATCTGGCCACTGCTGACACAGTTA	3424
Qy	2060	TCATTTATGACAGTGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC	2119
Db	3425	TTATCTATGACTCTGACTGGAAACCCCATTAATGACATTCAGGCCCTTTAGCAGAGCTCACC	3484
Qy	2120	GACTTGGCCAAAACAAATAAGGTGATGATTTATAGGCTCATAAACCGAGGCCACCATTTGAAG	2179
Db	3485	GGATTGGCAAAATAAAAAGGTAATGATCTACCGGTTTGTGACCCGTGCGTCAGTGGAGG	3544
Qy	2180	AAAGGATGATGCAATTGACTAAAAAGAAAATGGTCTTAGAGCATCTTGTGTTGG-----	2234
Db	3545	AGCGCATCAGCAGGTGGCAAGAAAGAAAATGATGCTGACGCATCTAGTGGTGGCGCCTG	3604
Qy	2235	-----GAAACTCAAAACACAAAAACATTAATCAGGAAGAGTTTAGATGACATCATCAGGTATG	2290



Db 3605 GGCTGGGCTCCAAGACTGGATCTATGTCCAAACAGGAGCTTGATGATATCTCTCAAATTTG 3664

QY 2291 GATCAAAAGGAGCTTTTGTAGTGAAGATGATGAAGCAGGAAAGTCTGGAAAAATTCATT 2350

Db 3665 GCACTGAGGAACCTATTCAAGGATGAAGCCACTGATGGAGGAGAGACAAAGAGGGAG 3724

QY 2351 ATGATGATCGGCTATAGACAAATTTGCTTGA-----TCGTGATCTCTGTGGAGGCAGAGG 2404

Db 3725 AAGATAGCAGTGTATCCACTACGATGATAGGCCATTGAACGGCTGTAGACCGTAACC 3784

QY 2405 AAGTCTCAGTGGATGAAGAGGAGGAATGGATTCTTTAAAGGCTTTCAAGGTGGCTAATT 2464

Db 3785 AGGATGAGACTGAAGACACAGAATTGCAGGGCATGAATGAATATTTGAGCTCATTCAAAG 3844

QY 2465 TTGAATATATAGATGAATAATGAGGCAGCAGCATTAGAGGCACAGAGAGTCGCTGTGAAA 2524

Db 3845 TGGCCCAAGTATGTGTACGGGAAGAAGAAATGGGGGAGGAAGAGGATAGAACGGGAAA 3904

QY 2525 GCAAATCTTCAGCAGGCAATTCTGATAGAGCAAGTTATTGGGAAGAGTTGTTAAAAAGATA 2584

Db 3905 TCATTAAACAGGAAGAAAGTGTGGATCCCTGA---CTACTGGGAGAAATTTGCTCGGCACC 3961

QY 2585 AATTTGAGCTGCACGAGCTGAGGAGCTTAAATGCTCTTGAAAAAGGAAGAGAAAGTCGCA 2644

Db 3962 ATTATGAGCAGCAGCAAGAAGATCTAGCCCCGAAATCTGGGCAAGGAAAAAGAAATCCGTA 4021

QY 2645 AGCAGTTGGTATCCATTGAAG--AAGATGATCTTGCTGGTTTGGAGATGTGAGCTCTG 2701

Db 4022 AACAGGTCAACTACAATGATGGCTCCAGGAGGACCGAGATTGGCAGGACGACCACTCG 4081

QY 2702 ATGGAGATGAAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGACAAGGAGTTT 2761

Db 4082 ACAACCACTCCGATTACTCAGTGGCTTCAGAGGAAGGTGATGAAGACTTTGATGAACGTT 4141

QY 2762 AGACGGTTCGACGGCCGTA CAGAAGAAAGGTGCGCGATAAATTTGGAAACCAA----- 2812

Db 4142 CAGAAAGCTCCCCGTAGGCCCCAGTCGTAAAGGCCCTGCGGAATGATAAAGATAAGCCATTGC 4201

QY 2813 CTCGGTTGATGGAAGGTGAGGGGAGATCTTTACAGAGTACTGGGTTTCAACCAGAGTCAAA 2872

Db 4202 CTCCTCTGTTGGCCCGTGTGGTGGGAATATTGAAGTACTTGGTTTTAATGCTCGTCAGC 4261

QY 2873 GGGCCATTTTGTACAGACTTTTGATGAGGTATGG-----AGCTGGCAATTTTGATT 2923

Db 4262 GAAAAGCCTTTCTTAATGCAATTATGCGATATGTTATGCCACCTCAGGATGCTTTTACTA 4321

QY 2924 GGAAGGAGTTTGTCTCGCTTAAAGCAGAGAACCTTTGAAGAAATAAATGAATATGGAA 2983

Db 4322 CCCAGTGGCTTGTAAAGACCTGCGAGGCAATCAGAGAAAGATTCAAGGCATATGTCT 4381

QY 2984 TACTCTTCTTGAAGCACATTGCTGAAGAAATAGACGAGAAATCTCCAACTTTTCAGATG 3043

Db 4382 CTCTTTTCATCGGCAATTATGTAGCCGGGGCAGATGGGGCTGAGACCTTTGCTGATG 4441

QY 3044 GTGTGCCCAAGGAGACTTAGAATAGAAATGTTCTAGTCAGAAATTGCTCTTCTGTATAC 3103

Db 4442 GTGTCCCCCGAAGGCTGTCTCGCCAGCATGTCTTACTAGAAATTGGTGTATGTCTT 4501

QY 3104 TAGTTCAGGAGAAAGGTGAAAATTTGTAGAAGATCATCCAGGG 3144

Db 4502 TGATTCGCAAGAAGGTTACGGAGTTTGAACATGTTAATGGG 4542

RESULT 10

AR269336

LOCUS AR269336 6328 bp DNA linear PAT 10-APR-2003

DEFINITION Sequence 1 from patent US 6500923.

ACCESSION AR269336

VERSION AR269336.1 GI:29700453

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

Unclassified.

REFERENCE 1 (bases 1 to 6328)

AUTHORS Seelig,H.P. and Renz,M.

TITLE Dermatomyositis-specific auto-antigen

JOURNAL Patent: US 6500923-A 1 31-DEC-2002;

FEATURES Location/Qualifiers

source 1..6328

/organism="unknown"

/mol\_type="genomic DNA"

ORIGIN

Query Match 10.1%; Score 421.4; DB 6; Length 6328;

Best Local Similarity 51.9%; Pred.No. 4.4e-88;

Matches 1236; Conservative 0; Mismatches 1076; Indels 69; Gaps 10;

QY 818 TACATCCATACCAGCTTGAGGAGCTTAATTTTTTTCGGGTTCTCGTGGTCAAAACAGACGC 877

Db 2177 TGCACCCCTATCAAATGGAGGCCCTGAATTGGTTGGCTTCTCCTGGGCTCAGGGCACTG 2236

QY 878 ATGTAATCCTTGTCTGATGAAATGGACTAGGCAAGACAATTCAAAGCATTTGCCCTTTTAG 937

Db 2237 ACACCATCTTGGCTGATGAGATGGGCTTGGGAAAACTGTACAGACAGCAGTCTTCTCTGT 2296

QY 938 CTTCACATTTTGGAGGAG-----AACCTCATTCGCGAATTTGGFAATTTGCTCTCTATCGA 991

Db 2297 ATTCCCTTTACAAGGAGGGTCAATTCAAAAGGCCCTTCTCCTAGTGAGCGGCCCTCTTTCTA 2356

QY 992 CTCTGCGTAACCTGGGAGAGAGATTGCCCACATGGGCCCCACAGATGAACGTGGTTATGT 1051

Db 2357 CCATCATCAACTGGAGCGGGAGTTTGAATGTGGCTCCAGACATGTATGTCGTAACCT 2416

QY 1052 ATTTTGGCACTGCGCAAGCTCGAGCAGTTATCAGAGAAACATGAGTTTACTTATCGAAAG 1111

Db 2417 ATGTGGGTGACAAGGACAGCCGTGCCATCATCCGAGAGAAATGAGTCTCCTTTGAAGACA 2476

QY 1112 ATCAAAAAAAGATCAAGAAAAAGAAATCTGGACAAATAAGTAGCGAAAGCAAGCAAAAA 1171

Db 2477 ATGCCATTCGTGGTGGCAAGAGGCCCTC-----CCGCGATGAAGAAAGAGGCAT 2524

QY 1172 GAATCAAGTTTGATGTCTCTCCTCACATCGTATGAGATGATCAACCTAGATTTCAGCAGTTC 1231

Db 2525 CTGTGAAATTCATGTCTGTGACATCCTATGAATTGATCACCATTGACATGGCTATTT 2584

QY 1232 TAAACCAATTAAGTGGGAGTGCAATGATTTGTATGAAGGTCAATCGACTGAAAAATAAGG 1291

Db 2585 TGGGCTCTATTGATTGGGCCTGCCTCATCGTGGATGAAGCCCATCGGCTGAAGAACAAATC 2644

QY 1292 ATTCAAGCTGTTCTCTTCATTGACACAGTATTCAAGTAACCCACCTGATTTCTTCTGACAG 1351

Db 2645 AGTCTAAGTCTTCCGGGTATTGAATGGTTACTCCTCCAGCACAAAGCTGTTGCTGACTG 2704

QY 1352 GAACACCACTTCAGAACCAACTTGGATGAACCTTTTCATGCTCATGCAATTTTCTTGATCGCG 1411

Db 2705 GGACACCAATTACAAAACAATCTGGAAGAGTTGTTTCATCTGCTCAACTTCTCACCCCG 2764

QY 1412 GGAAGTTTGGAAAGTTTGGAGAGTTCAGGAGGAGTTCAAAGATAATTAATCAAGAGGAGC 1471

Db 2765 AGAGGTTCCACAATTTGGAAGGTTTTTTTGGAGGAGTTTGTGCTGACATTTGCCAAGGAGACC 2824

QY 1472 AGATCTCAAGGTTGCACAAAAATGTTGGCTCCACATTTGCTCAGAAAGGTAAAAAAGACG 1531

Db 2825 AGATAAAAAAAACTGCATGACATGCTGGGGCCGCACATGTTGCGGCGGCTCAAAAGCCGATG 2884

QY 1532 TAATGAAAGACATGCCCCCAAAAAAGGAGAGCTCATTTTTCGTTGTTGATCTGAGCAGTCTGC 1591

Db 2885 TGTTCAGAAACATGCCCTCCAAGACAGAACTAATTGTGCGTGTGGAGCTGAGCCCTATGC 2944

QY 1592 AGAAAGAAATATACAAAGCTATTTTACCCGTAATTATCAAGTATTGA--CAAAAAAGG 1648

Db 2945 AGAAGAAATACTACAAGTACATCTCTCACTCGAAATTTTGAAGCACTCAATGCCCGAGGTG 3004

QY 1649 GAGGTGCTCAAAATTTCCCTTAATAACATTATGATGGAATTACGAAAGATATGCTGCCATC 1708

Db 3005 GTGGCAACCAAGGTGTCTCTGTCTGAATGTGGTATGGATCTTTAAGAAAGTGTGCAACCATC 3064







VERSION AX332238.1 GI:18122872
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
Horrigan,S., Soppet,D.R. and Weaver,Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 2747 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
Location/Qualifiers
source 1..6417
/organism="Homo sapiens"
/mol\_type="unassigned DNA"
/db\_xref="taxon:9606"
ORIGIN
Query Match 10.1%; Score 421.4; DB 6; Length 6417;
Best Local Similarity 51.9%; Pred. No. 4.4e-88;
Matches 1236; Conservative 0; Mismatches 1076; Indels 69; Gaps 10;
QY 818 TACATCCATACCAGCTTGAGGGACTTAATTTTTCGCGTTCTCGTGGTCAAAACAGACGC 877
Dd 2266 TGCACCCCTATCAATGGAGGGCCCTGAATTGGTTGCGCTTCTCTGGGCTCAGGGCAGTG 2325
QY 878 ATGTAATCCTTGCTGATGAAATGGGACTAGGCAAGACAATTCAAAGCATTTGCCCTTTTAG 937
Dd 2326 ACACCATCTTGGCTGATGAGATGGGCTTGGGAAAACGTGTACAGACAGCAGTCTTCCTGT 2385
QY 938 CTTCACTTTTGGAGGAG-----AACCTCATTCGCAATTTGGTAATTGCTCCTCTATCGA 991
Dd 2386 ATTCCCTTTACAAGGAGGGTTCATTCCAAAGGCCCTTCCCTAGTGAGCGCCCTCTTTCTA 2445
QY 992 CTCTGCGTAACCTGGGAGAGAGATTGGCCACATGGGCCCCACAGATGAAGCGTGGTTATGT 1051
Dd 2446 CCATCATCAACTGGGAGCGGAGTTTGAATGTGGGCTCCAGACATGTATGTCGTAACCT 2505
QY 1052 ATTTTGGCACTGCGCAAGCTCGAGCAGTTATCAGAGAACATGAGTTTACTTATCGAAAG 1111
Dd 2506 ATGTGGGTGACAAGGACAGCCGTGCCATCATCCGAGAGAATGAGTTCTCCTTTGAAGACA 2565
QY 1112 ATCAAAAAAAGATCAAGAAAAAGAAATCTGGACAAATAAGTAGCGAAAGCAAGCAAAAA 1171
Dd 2566 ATGCCATTCGTGGTGCAAGAGGCCCTC-----CCGCATGAAGAAAGAGGCAT 2613
QY 1172 GAATCAAGTTTGATGTCTCTCTACATCGTATGAGATGATCAACCTAGATTCAGCAGTTTC 1231
Dd 2614 CTGTGAATTCATGTGCTGTGACATCCTATGAATTGATCACCATTGACATGGCTATTT 2673
QY 1232 TAAACCAATTAAGTGGGAGTGATGATGTTGATGAAGGTCTACGACTGAAATAAAGG 1291
Dd 2674 TGGGCTCTATTGATTGGGCCCTGCCTCATCGTGGATGAAGCCCATCGGCTGAAGAACATC 2733
QY 1292 ATTCAAAGCTGTTCTCTTCATTGACACAGTATTCAAGTAACCAACCGTATTCTTCTGACAG 1351
Dd 2734 AGTCTAAGTTCTTCGGGTATTGAATGGTTACTCACTCCAGACAAGCTGTTGCTGACTG 2793
QY 1352 GAACACCATTTCAGAACAACTTGGATGAACCTTTTCATGCTCATGCTATGCATTTTCTTGATGCGG 1411
Dd 2794 GGACACCATTACAAAACAATCTGGAAGAGTTGTTTCATCTGCTCAACTTTCTCACCCCCG 2853
QY 1412 GGAAGTTTGAAGTTTGGAGGAGTTCCAGGAGGAGTTTCAAAGATATTAATCAAGAGGAGC 1471
Dd 2854 AGAGGTTCCACAAATTTGGAAGGTTTGTGGAGGAGTTTGTGACATTTGCCAAGGAGACC 2913
QY 1472 AGATCTCAAGGTGACAAAAATGTTGGCTCCACATTTTGCTCAGAAGGGTAAAAAAGACG 1531
Dd 2914 AGATAAAAAAATGCAATGATGCTGGGCCCGCACATGTTGCGGGCGGCTCAAAGCCGATG 2973
QY 1532 TAATGAAGACATGCCCCCCCCAAAAAGGAGCTCATTTTTCGTGTGATCTGAGCAGTCTGC 1591

Db 2974 TGTTCAAGAACATGCCCTCCAAGACAGAACTAATTGTGCGTGTGGAGCTGAGCCCTATGC 3033
QY 1592 AGAAAGAATATTACAAAGCTATTTTACCCGTAATTATCAAGTATTGA---CAAAAAAGG 1648
Dd 3034 AGAAGAAATACTACAAGTACATCCTCACTCGAAATTTTGAAGCACTCAATGCCGAGGTG 3093
QY 1649 GAGTGTCTCAAAATTTCCCTTAATAAACATTATGATGGAATTACGAAAAAGTATGCTGCCATC 1708
Dd 3094 GTGGCAACCAGGTGTCTCTGCTGAATGTGGTATGATCTTAAGAAGTGTGCAACCATC 3153
QY 1709 CTTA-----TATGCTAGAGGGTGTGAGCCAGTTATTACGACGCAAAATGAAGCTT 1759
Dd 3154 CATACCTCTTCCCTGTGGCTGCAATGGAAGCTCCTAAGATGCCTAATGGCATGTATGATG 3213
QY 1760 TCAAAACAATTTTGGAGTCTTGTGAAAAAGCTGCAACTTCTAGATAAAAATGATGGTCAAAAC 1819
Dd 3214 GCAGTGGCCCTAATCAGAGCATCTGGGAAATATTATTGCTGCTGCAGAAAAATGCTCAAGAACC 3273
QY 1820 TGAAGAGCAAGGACACAGAGTCTCTAATATACACACAGTTTCAGCATATGCTGGACTTAC 1879
Dd 3274 TTAAGGAGGGTGGGCATCGTGTACTCATCTTTTCCCAGATGACCAAGATGCTAGACCTGC 3333
QY 1880 TTGAAGACTACTGTACCCTAAGAAATGGCAGTACGAGCGCAATTGATGGAAAGGTTGGCG 1939
Dd 3334 TAGAGGATTTCTTGGAAACATGAAGGTTATAAATACGAACGCATCGATGGTGAATCACTG 3393
QY 1940 GAGCTGAGCGGCAAAATACGCATAGATCGGTTCAATGCCAAAAATTTCTAAACAAGTTTGT 1999
Dd 3394 GGAACATGCGGCAAGAGGCCATTGACCGCTTCAATGCACCGGTGCTCAGCAGTTCTGCT 3453
QY 2000 TTTTGTCTCTCCACAAGAGCTGGTGGCTTAGGAATAAATCTTGCAACGGCTGATACAGTAA 2059
Dd 3454 TCTTGTCTTCCACTCGAGCTGGGGCCCTTGAATCAATCTGGCCACTGCTGACACAGTTA 3513
QY 2060 TCATTTATGACAGTGACTGGAATCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC 2119
Dd 3514 TTATCTATGACTCTGACTGGAAACCCCATTAATGACATTCAGGCTTTTAGCAGAGCTCAC 3573
QY 2120 GACTTGGCCAAAACAAATAAGGTGATGATTTATAGGCTATAAACCGAGGCACCATTTGAAG 2179
Dd 3574 GGATTTGGCAAAATAAAAAAGGTAATGATCTACCGGTTTGTGACCCGTCGTCAGTGGAGG 3633
QY 2180 AAAGGATGATGCAATTGACTAAAAAGAAAAATGGTTCTAGAGCATCTTGTGTTGG----- 2234
Dd 3634 AGCGCATCAGCAGGTGGCAAGAGAAATGATGCTGACGCATCTAGTGGTGGCGCCTG 3693
QY 2235 ----GAAACTCAAAACACAAAACATTAAATCAGGAAGAGTTAGATGACATCATCAGGTATG 2290
Dd 3694 GGGTGGCTCCAAGACTGGATCTATGTCCAAACAGGAGCTTGATGATATCCTCAAATTTG 3753
QY 2291 GATCAAAAGGAGCTTTTGTAGTGAAGATGATGAAGCAGGAAAAGTCTGGAATAATTCATT 2350
Dd 3754 GCACTGAGGAACTATTCAAGGATGAAGCCACTGATGGAGGAGAGACAAACAAGAGGGAG 3813
QY 2351 ATGATGATCGGCTATAGACAAATTCCTTGA-----TCGTGATCTCGTGGAGGCAGAGG 2404
Dd 3814 AAGATAGCAGTGTATTCCTACTACGATGATAAGGCCATTGAACGGCTGTAGACCGTAACC 3873
QY 2405 AAGTCTCAGTGGATGATGAAGAGGAGAATGGATTCTTAAAGGCTTCAAGGTGGCTAATT 2464
Dd 3874 AGGATGAGACTGAAGACACAGAAATTCAGGGCATGAATGAATAATTGAGCTCATTCAAAG 3933
QY 2465 TTGAATATATAGATGAAATAGAGGCAGCAGCATTTAGAGGCACAGAGAGTCGCTGCTGAAA 2524
Dd 3934 TGGCCAGTATGTGGTACGGGAAGAAGAAATGGGGAGGAAGAGGAGGTAGAACGGGAAA 3993
QY 2525 GCAATCTTCAGCAGGCAATCTGATAGAGCAAGTTATTGGGAAGAGTTGTTAAAAAGATA 2584
Dd 3994 TCATTAAACAGGAAGAAAGTGTGGATCCTGA---CTACTGGGAGAAATTTGCTGCGGCACC 4050
QY 2585 AATTGAGCTGCACCAGGCTGAGGAGCTTAATGCTCTTGGAAAAAGGAAGAGATCGCA 2644



QY 1172 GAATCAAGTTTGATGTCCTCCTCACATCGTATGAGATGATCAACCTAGATTGAGAGTTC 1231  
Db 2614 CTGTGAATTCATGCTGCTGACATCCTATGAATTGATCACCATTGACATGGCTATTT 2673  
QY 1232 TAAACCAATTAAGTGGGAGTGCATGATGTTGATGAAGGTTCATCGACTGAAAAATAAGG 1291  
Db 2674 TGGGCTCTATTGATTGGGCTGCCCTCATCTGTGATGAAGCCCATCGGCTGAAGAACATC 2733  
QY 1292 ATTCAAAGCTGTTCTCTTATTGACACACAGTATTCAAGTAACCAACCGTATTTCTTGACAG 1351  
Db 2734 AGTCTAAGTTCTTCGGGTATTGAATGGTTACTCACTCCAGCACAAAGCTGTTGCTGACTG 2793  
QY 1352 GAACACCACCTTCAGAAACAACCTTGGATGAACCTTTTCATGCTCATGCTATTTCTTGATGCGG 1411  
Db 2794 GGACACCATTACAAAAACAATCTGGAAGAGTTGTTTCATCTGCTCAACTTCTCACCCCGC 2853  
QY 1412 GGAAGTTTGAAGTTTGGAGGAGTTCAGGAGGAGTTCAGGAGGAGTTCAAAGATATTAATCAAGAGGAGC 1471  
Db 2854 AGAGGTTCCACAATTTGGAAGGTTTTTTGGAGGAGTTTGTGTCATGATTCGCAAGGAGGACC 2913  
QY 1472 AGATCTCAAGTTGACAAAAAATGTTGGCTCCACATTTTGCTCAGAAAGGTAATAAAGACG 1531  
Db 2914 AGATAAAAAAATGACATGCTGGGCGGCACATGTTGGGCGGCTCAAAGCCGATG 2973  
QY 1532 TAATGAAAGACATGCCCCCAAAAAAGGAGTCAATTTTGGGTTGATCTGAGCAGTCTGC 1591  
Db 2974 TGTTCAAGAACATGCCCTCCAAGACAGAACTAATTTGTCGCTGAGGAGTGGCCCTATGC 3033  
QY 1592 AGAAAGAAATATCAAAAGCTATTTTACCCGTAATTAATCAAGTATGA---CAAAAAAGG 1648  
Db 3034 AGAAGAAATACTACAAAGTACATCCTCACTCGAAATTTTGAAGCACTCAATGCCGAGGTG 3093  
QY 1649 GAGGTGCTCAAAATTTCCCTTAATAACATTAATGATGGAATTACGAAAGATGCTGCCATC 1708  
Db 3094 GTGGCAACCAGGTGCTCTGCTGAATGTGTGATGGATCTTAAGAAAGTGTGCAACCATC 3153  
QY 1709 CTTA-----TATGCTAGAGGGTGTGAGCCAGTTATTACAGCGCAAAATGAAGCTT 1759  
Db 3154 CATACCTCTTCCCTGTGGCTGCAATGGAAGCTCCTAAGATGCCTAATGGCATGTATGATG 3213  
QY 1760 TCAAAACAACCTTTTGGAGTCTTGTGGAAGCTGCAACTTCTAGATAAAATGATGTCAAAC 1819  
Db 3214 GCAGTGCCCTAATCAGAGCATCTGGGAATATTGCTGCTGCAGAAATGCTCAAGAAC 3273  
QY 1820 TGAAGAGCAAGGACACAGAGTCCCTAATATACACACAGTTTTCAGCATATGCTGGACTTAC 1879  
Db 3274 TTAAGGAGGTGGGCATCGTGTACTCATCTTTTCCAGATGACCAAGATGCTAGACCTGC 3333  
QY 1880 TTGAAGACTACTGTACCCATAAGAAATGGCAGTACGAGCGAATTGATGGAAAGTTGGCG 1939  
Db 3334 TAGAGGATTTCTTGGAACATGAAGGTTATPAAATACGAACGCATCGATGGTGAATCACTG 3393  
QY 1940 GAGCTGAGCGGCAAAATACGCATAGATCGGTTCAATGCCAAAAATCTPAAACAAGTTTGT 1999  
Db 3394 GGAACATGCGGCAAGAGGCCATTGACCGCTTCAATGCACCGGTGCTCAGCAGTTCTGCT 3453  
QY 2000 TTTTGTCTCCACAAGAGCTGGTGGCTTAGGAATAAATCTTGCAACGGCTGATACAGTAA 2059  
Db 3454 TCTTGTCTTCCACTCGAGCTGGGGCCCTTGGAAATCAATCTGGCCACTGCTGACACAGTTA 3513  
QY 2060 TCATTTATGACAGTGACTGGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC 2119  
Db 3514 TTATCTATGACTCTGACTGGAAACCCCAATAATGACATTCAGGCCCTTTAGCAGAGCTCAC 3573  
QY 2120 GACTTGGCCAAACAAATAAGGTGATGATTTATAGGCTCATAAACCGAGGCACCATTTGAAG 2179  
Db 3574 GGATTGGGCAAAATAAAAAAGGTAATGATCTACCGGTTTGTGACCCGTGCTGAGTGAGG 3633  
QY 2180 AAAGGATGATGCAATTTGACTAAAAAGAAATGGTTCTAGAGCATCTTGTGTTGG----- 2234  
Db 3634 AGCGCATCACGAGGTGGCAAGAAGAAATGATGCTGACGCCATCTAGTGGTGGCGCCTG 3693  
QY 2235 ----GAAACTCAAAACACAAAAACATTTAATCAGGAAGAGTTAGATGACATCATCAGGTATG 2290

Db 3694 GGCTGGGCTCCAAGACTGGATCTATGTGTCCAAACAGGAGCTTGTATGATATCTCAAATTTG 3753  
QY 2291 GATCAAGGAGCTTTTGTGCTAGTGAAGATGATGAAGCAGGAAAGTCTGGAAAAATTCATT 2350  
Db 3754 GCACTGAGGAACCTATTCAAGGATGAAGCCACTGATGGAGGAGAGACAACAAAGAGGGAG 3813  
QY 2351 ATGATGATGCGGCTATAGACAAATTGCTTGA-----TCGTGATCTCGTGGAGGCAGAGG 2404  
Db 3814 AAGATAGCAGTGTATTCCACTACGATGATAAGGCCATTGAACGGCTGTAGACCGTAACC 3873  
QY 2405 AAGTCTCAGTGGATGATGAAGAGGAGAAATGGATTCTTAAAGGCTTTCAAGGTGGTAATT 2464  
Db 3874 AGGATGAGACTGAAGACACAGAAATTCAGGGCATGAATGAATATTGAGCTCATTCAAAG 3933  
QY 2465 TTGAATATATAGATGAAAAATGAGCAGCAGCATTTAGAGGCACAGAGAGTCTGCTGAAA 2524  
Db 3934 TGGCCCATGATGTGGTACGGGAAGAAATGGGGAGGAAGAGGAGGTAGAACGGGAAA 3993  
QY 2525 GCAATCTTCAGCAGGCAATTCTGATAGAGCAAGTTATTGGGAAGAGTTGTTAAAAAGATA 2584  
Db 3994 TCATTAAACAGGAAGAAAGTGTGATCCTGA---CTACTGGAGAAATTTGCTGCGGCACC 4050  
QY 2585 AATTGAGCTGCACCAGGCTGAGGAGCTTAATGCTCTTGTGTTTGGAAAGATGTGAGCTCTG 2701  
Db 4111 AACAGGTCAACTACAATGATGGCTCCCAGGAGGACCGAGATTGSCAGGACGACCATCCG 4170  
QY 2702 ATGGAGATGAAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGACAAGGAGTTC 2761  
Db 4171 ACAACCAGTCCGATTACTCAGTGGCTTCAGAGGAAGGTGATGAAGACTTTGATGAACGTT 4230  
QY 2762 AGACGGGTGACGGCCGTACAGAAGAAAGGTCGCGATAATTTGGAAACCAA----- 2812  
Db 4231 CAGAAGCTCCCCGTAGGCCCAAGTCGTAAGGGCCCTGCGGAATGATAAAGATAAGCCATTGC 4290  
QY 2813 CTCGGTTGATGGAAGGTGAGGGAGATCTTTTTCAGAGTACTGGGTTTCAACCAGAGTCAAA 2872  
Db 4291 CTCCTCTGTTGGCCCGTGTGTTGGGGAATATTGAAGTACTTGGTTTAAATGCTCGTCAGC 4350  
QY 2873 GGGCCATTTTGTACAGACTTTTGTGATGAGGTATGG-----AGCTGGCAATTTTGATT 2923  
Db 4351 GAAAAGCCTTTCTTAATGCAATATATCGATATGGTATGCCACCTCAGGATGCTTTTACTA 4410  
QY 2924 GGAAGGAGTTTGTTCCTCGCTTAAAGCAGAAGACCTTTTGAAGAAATAAATGAATATGAA 2983  
Db 4411 CCCAGTGGCTTGTAAAGAGACCTCGCAGGCAAAATCAGAGAAAGAGTTCAAGGCATATGCT 4470  
QY 2984 TACTCTTCTTGAAGCACATTTGCTGAAGAAATAGACGAGAAATTTCTCAACCTTTTCAGATG 3043  
Db 4471 CTCCTTTTCATGCGGCATTTTATGTGAGCCGGGGCAGATGGGGCTGAGACCTTTGCTGATG 4530  
QY 3044 GTGCCCCAAGGAAGGACTTGAATAGAAAGATGTTCTAGTCAAGAAATGCTCTTCTGTATAC 3103  
Db 4531 GTGTCCCCCGAGAAGGCTGTCTCGCCAGCATGTCTTACTAGAAATGGTGTATGTCTT 4590  
QY 3104 TAGTTCAAGGAAGGTGAAATTTTGTAGAAGATCATCCAGG 3144  
Db 4591 TGAITCGCAAGAAGGTTTCAGGAGTTTGAACATGTTAATGGG 4631

RESULT 14  
AR338834  
LOCUS AR338834 6475 bp DNA linear PAT 17-AUG-2003  
DEFINITION Sequence 325 from patent US 6569662.  
ACCESSION AR338834  
VERSION AR338834.1 GI:33725691  
KEYWORDS Unknown.  
SOURCE Unknown.  
ORGANISM Unknown.



Unclassified.									
REFERENCE	1	(bases 1 to 6475)							
AUTHORS	Tang,Y.T., Zhou,P. and Drmanac,R.T.								
TITLE	Nucleic acids and polypeptides								
JOURNAL	Patent: US 6569662-A 325 27-MAY-2003;								
FEATURES	Location/Qualifiers								
	source	1..6475							
	/organism="unknown"								
ORIGIN									
Query Match 10.1%; Score 421.4; DB 6; Length 6475;									
Best Local Similarity 51.9%; Pred. No. 4.4e-88;									
Matches 1236; Conservative 0; Mismatches 1076; Indels 69; Gaps 10;									
Qy	818	TACATCCATACCAAGCTTGAGGGACTTAATTTTTCGGTTCTCGTGGTCAAAACAGACGC	877						
Db	2395	TGCACCCCTATCAAAATGGAGGGCCTGAATTGGTTGCGCTTCTCCTGGCTCAGGGCACTG	2454						
Qy	878	ATGTAATCCTTGCTGATGAAATGGGACTAGGCAAGACAAATTCAAAGCATTCGCCCTTTTAG	937						
Db	2455	ACACCATCTTGGCTGATGAGATGGGCCCTTGGGAAAACGTACAGACAGCAGTCTTCCTGT	2514						
Qy	938	CTTCACITTTTGGAGGAG-----AACCTCATTCGCGCATTTGGTAATTGCTCCTCTATCGA	991						
Db	2515	ATTCCCTTTACAAGGAGGGTCATTCCAAAGGCCCTTCTTAGTGAGCGCCCTCTTTCTA	2574						
Qy	992	CTCTGCGTAACCTGGGAGAGAGAGTTTGGCCACATGGGCCCCACAGATGAACGTGGTTATGT	1051						
Db	2575	CCATCATCAACTGGGAGCGGAGTTTGAATGTGGGCTCCAGACATGTATGCTAACCT	2634						
Qy	1052	ATTTTGGCACTGCGCAAGCTCGAGCAGTTATCAGAGAACATGAGTTTACTTATCGAAAG	1111						
Db	2635	ATGTGGGTGACAAGGACAGCCGTGCCATCATCCGAGAGAATGAGTTCTCCTTTGAAGACA	2694						
Qy	1112	ATCAAAAAAGATCAAGAAAAAGAAATCTGGACAAATAAGTAGCGAAAGCAAGCAAAAA	1171						
Db	2695	ATGCCATTCTGTGGTGGCAAGAGGCCTC-----CCGCATGAAGAAAGAGGCAT	2742						
Qy	1172	GAATCAAGTTTGATGTCCTCCTCACATCGTATGAGATGATCAACCTAGATTGAGCAGTTC	1231						
Db	2743	CTGTGAAATTCCATGTGCTGCTGACATCCTATGAATTGATCACCATTGACATGGCTATTT	2802						
Qy	1232	TAAACCAATTAAAGTGGGAGTGCAATGATTGTTTCATGAAGGTCAATCGACTGAAAAATAAGG	1291						
Db	2803	TGGGCTCTATTGATTGGGCCTGCCTCATCGTGGATGAAGCCCATCGGCTGAAGAAACAATC	2862						
Qy	1292	ATTCAAAGCTGTTCTCTTCATTGACACAGTAATCAAGTAACCAACCGTATTCCTCTGACAG	1351						
Db	2863	AGTCTAAGTTCTTCGGGTATTGAATGGTTACTCACTCCAGCACAAAGCTGTTGCTGACTG	2922						
Qy	1352	GAACACCACCTTCAGAAACAACTTGGATGAACITTTTCATGCTCATGCAATTTCTTGATCGG	1411						
Db	2923	GGACACCATTACAAAAACAATCTGGAAGAGTTGTTTCATCTGCTCAACTTCTCACCCCCG	2982						
Qy	1412	GGAAGTTTGGAGTTTGGAGGAGTTCAGGAGGAGTTCCAGGAGGAGTTCAAAGATATTAATCAAGAGGAGC	1471						
Db	2983	AGAGGTTCCACAAATTTGGAAGGTTTTTTGGAGGAGTTTGCTGACATTTGCCAAGGAGGACC	3042						
Qy	1472	AGATCTCAAGTTTGACAAAAATGTTGGCTCCAATTTGCTCAGAAAGGTAAAAAAGACG	1531						
Db	3043	AGATAAAAAAACTGCATGACATGCTGGGCCGCGACATGTTGCGGGGCTCAAAGCCGATG	3102						
Qy	1532	TAATGAAAGACATGCCCCCCCAAAAAGGAGCTCAATTTTTCGCTGTTGATCTGAGCAGTCTGC	1591						
Db	3103	TGTTCAAGAACAATGCCCTCCAAGACAGAACTAATTGTGCGTGTGGAGCTGAGCCCTATGC	3162						
Qy	1592	AGAAAGAATATTACAAAGCTATTTTACCCGTAATTATCAAGTATTGA---CAAAAAAGG	1648						
Db	3163	AGAAGAAATACTACAAGTACATCCTCACTCGAAATTTTGAAGCACTCAATGCCCGAGGTG	3222						
Qy	1649	GAGGTGCTCAAATTTCCCTTAATAACATTTATGATGGAATTACGAAAAAGTATGCTGCCATC	1708						

Db	3223	GTGGCAACCAAGGTGTCTCTGCTGAATGTGGTGATGGATCTTTAAGAAGTGCTGCAACCATC	3282						
QY	1709	CTTA- - - - - TATGCTAGAGGGTGTGAGCCAGTTATTACGACGCAAAATGAAGCTT	1759						
Db	3283	CATACCTCTTCCCTGTGGTGCATGGAAGCTCCTAAGATGCCTAATGGCATGTATGATG	3342						
QY	1760	TCAAACAACCTTTTGGAGTCTTGTGGAAGCTGCAACTTCTAGATAAAAATGATGTCAAAAC	1819						
Db	3343	GCAGTGCCCTAATCAGAGCATCTGGGAAATATTGCTGCTGCAGAAAAATGCTCAAGAACC	3402						
QY	1820	TGAAAGAGCAAGGACACAGAGTCTCTAATATATACACACAGTTTTCAGCATATGCTGGACTTAC	1879						
Db	3403	TTAAGGAGGTGGGCATCGTGTACTCATCTTTTCCCAGATGACCAAGATGCTAGACCTGC	3462						
QY	1880	TTGAAGACTACTGTACCCATAGAAATGGCAGTACGAGCGAATTGATGGAAAGGTTGGCG	1939						
Db	3463	TAGAGGATTTCTTGGAACATGAAGTTATAAATACGAACGCATCGATGGTGAATCACTG	3522						
QY	1940	GAGCTGAGCGGCAAAATACGCATAGATCGGTTCAATGCCAAAAATTCTAAACAAGTTTGT	1999						
Db	3523	GGAACATGCGGCAAGAGGCCATTGACCGCTTCAATGCACCGGTGCTCAGCAGTTCTGCT	3582						
QY	2000	TTTTGCTCTCCACAAGAGCTGGTGGCTTAGGAATAAATCTTGCAACGGCTGATACAGTAA	2059						
Db	3583	TCCTGCTTCCACTCGAGCTGGGCGCTTGAATCAATCTGCCCACTGCTGACACAGTTA	3642						
QY	2060	TCATTTATCAGAGTGACTGGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC	2119						
Db	3643	TTATCTATGACTCTGACTGGAACCCCCATAATGACATTCAGGCTTTTAGCAGAGCTCACC	3702						
QY	2120	GACTTGGCCAAAACAAATAAGGTGATGATTTATAGGCTCATAAACCGAGGCACCATTTGAAG	2179						
Db	3703	GGATTGGGCAAAATAAAAAGGTAATGATCTACCGGTTTGTGACCCGTGCGTCAGTGGAGG	3762						
QY	2180	AAAGGATGATGCAATTGACTAAAGAAAGAAATGGTCTAGAGCATCTTGTGTTGG- - - - -	2234						
Db	3763	AGCGCATCACGCGAGGTGGCAAGAAAGAAATGATGCTACGCACTAGTGTGGTGGCGCTG	3822						
QY	2235	- - - - - GAAACTCAAAAACAAAAACATTAATCAGGAAGAGTTAGATGACATCATCAGGTATG	2290						
Db	3823	GGCTGGGCTCCAAGACTGGATCTATGTCCAAAACAGGAGCTTGATGATATCCTCAAATTTG	3882						
QY	2291	GATCAAAGGAGCTTTTGTCTAGTGAAGATGATGAAGCAGGAAAAGTCTTGAAAAAATTCATT	2350						
Db	3883	GCACTGAGGAACTATTCAAGGATGAAGCCACTGATGGAGGAGGAGACAAACAAAGAGGAG	3942						
QY	2351	ATGATGATCGCGCTATAGACAAATTTGCTTGA- - - - - TCGTGATCTCGTGAGGCAGAGG	2404						
Db	3943	AAGATAGCAGTGTATCCACTACGATGATAAGGCCATTGAAACGGCTGTAGACCGTAACC	4002						
QY	2405	AAGTCTCAGTGGATGATGAAGAGGAGAAATGGATTCTTAAAGGCTTTCAGAGGTGGCTAATT	2464						
Db	4003	AGGATGAGACTGAAGACACAGAAATTGCAGGGCATGAATGAATATTGAGCTCATTCAAAG	4062						
QY	2465	TTGAATATATAGATGAAAAATGAGGCAGCAGCATTAGAGGCACAGAGTCGCTGTGAAA	2524						
Db	4063	TGGCCCAGTATGTGGTACGGGAAGAGAAATGGGGAGGAAAGAGGAGGTAGAACGGGAAA	4122						
QY	2525	GCAAACTCTTCAGCAGGCAATTCGTATAGAGCAAGTATTGGGAAGAGTTGTTAAAAAGATA	2584						
Db	4123	TCATTTAAACAGGAAGAAAGTGTGGATCCCTGA- - - CTACTGGGAGAAATGTCTGCGGCACC	4179						
QY	2585	AATTTGAGCTGCACCAGGCTGAGGAGCTTAATGCTCTTGGAAAAAGGAAGAGAGTCGCA	2644						
Db	4180	ATTATGAGCAGCAGCAAGAAGATCTAGCCCGAAATCTGGGCAAAAGGAAAAAAGATCCGTA	4239						
QY	2645	AGCAGTTGGTATCCATTGAAG- - - AAGATGATCTTGTGTTTGGAAAGATGTGAGCTCTG	2701						
Db	4240	AACAGGTCAACTACATGATGGCTCCAGGAGGACCGAGATTGGCAGGACGACCGAGTCCG	4299						
QY	2702	ATGGAGATGAAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGACAAGGAGTTC	2761						
Db	4300	ACAACCAAGTCCGATTACTCAGTGGCTTCAGAGGAAGGTGATGAAGACTTTTGATGAACGTT	4359						

Qy	2762	AGACGGGTGACGGCCGTACAGAAAGAGGGTCGCGATAAATTTGGAAACCAA	-----	2811
Db	4360	CAGAAAGCTCCCGTAGGCCCAAGTCGTAAGGGCTGCGGAATGATAAAGATAAGCCATTGC		4419
Qy	2813	CTCCGTTGATGGAAGGTGAGGGGAGATCTTTTCAGAGTACTGGGTTTCAACCAGAGTCAAA		2872
Db	4420	CTCCTCTGTTGGCCCGTGTGGTGGGAATAFTGAAGTACTTGGTTTTAATGCTCGTCAGC		4479
Qy	2873	GGGCCATTTTGTACAGACTTTTGATGAGGTATGG-----AGCTGGCAATTTTGATT		2923
Db	4480	GAAAAGCCTTCTTAATGCAATTAATGCGATATGGTATGCCACCTCAGGATGCTTTTACTA		4539
Qy	2924	GGAAGGAGTTTGTTCCTCGCTTAAAGCAGAAGACCTTTGAAGAAATAAAATGAATATGAA		2983
Db	4540	CCCAGTGGCTGTGAAGAGACCTGCGAGGCAAAATCAGAGAAAGAGTTCAAGGCATATGTCT		4599
Qy	2984	TACTCTTCTTGAAGCACATTGCTGAAGAAATAGACGAGAAATTTCTCCAACCTTTTCAGATG		3043
Db	4600	CTCTTTTCATGCGGCATTTATGTGAGCCGGGGCGAGATGGGGCTGAGACCTTTGCTGATG		4659
Qy	3044	GTGTGCCCCAAGGAAGGACTTAGAATAGAAGATGTTTCTAGTCAGAATTTGCTCTTCTGATAC		3103
Db	4660	GTGTCCCCCGAAGAGGCTGTCTCGCCAGCATGTCTTACTAGAATTTGGTGTATGTCTT		4719
Qy	3104	TAGTTCAGGAGAAGGTGAAATTTGTAGAAGATCATCCAGGG	3144	
Db	4720	TGATTCGCAAGAAGGTTTCAGGAGTTTGAACATGTTAATGGG	4760	
RESULT 15				
AX924075				
LOCUS				
AX924075				
DEFINITION				
AX924075				
ACCESSION				
AX924075.1				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
FEATURES				
source				
ORIGIN				
Query Match				
Best Local Similarity				
Matches 1233; Conservative				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				
Qy				
Db				

Search completed: September 18, 2004, 20:09:01  
Job time : 15797 secs

Qy	2112	AGCTCATCGACTTGGCCAAACAATAAAGGTGATGATTTATAGGCTCATAAACCGAGGCAC	2171
Db	3717	GGCTCATCGGATTGGCCAGGCCAACAAAGTGATGATTTACCGTTTGTGACTCGCGCGTC	3776
Qy	2172	CATTGAAGAAAGGATGATGCAATTGACTAAAAAGAAATGGTTCTAGAGCATCTTGTGTG	2231
Db	3777	AGTGAAGAGCGAATCACACAAGTGGCCAAGAGAAAGATGATGCTGACACACCTGGTTGT	3836
Qy	2232	TGG-----GAAACTCAAAACACAAACACATTAATCAGGAAGAGTTAGATGACATCAT	2282
Db	3837	GCGGCTGGGCTGGGCTCCAAGGCAGGCTCCATGTCCAAGCAGGAGCTTGACGACATCT	3896
Qy	2283	CAGGTATGGATCAAAGGAGCTTTTTCCTAGTGAAGATGATGAAGCAGGAAAGTCTG----	2338
Db	3897	CAAAATTGGCACTGAAGAGCTAATTCAGGATGAAAACGAGGGGAGAAACAAGGAGGAGA	3956
Qy	2339	-----GAAAAATTCAATTATGATGATGCGGCTATAGACAAATTGCTTGATCGTGATCTCG	2393
Db	3957	CAGCAGTGTGATTCATTATGACAAATGAGGCCATCGCTCGGCTGTGGACCGGAACCCAGGA	4016
Qy	2394	GGAGGCAGAGGAAGTCTCAGTGGATGATGAAGAGGAGGAAATGGATTCTTAAAGGCTTTCAA	2453
Db	4017	TGCAACTGAGGA-----CACTGACGTGCAGAACATGAATGAGTATCTCAGCTCCTTCNA	4070
Qy	2454	GGTGGCTAATTTGAATATATAGATGAATAATGAGGCAGCAGCATTAGAGGCACAGAGAGT	2513
Db	4071	GGTGGCACAGTACGTCGTGCGGGAAGAAGACAAAGATTGAGGAAATTGAGCGAGAGATCAT	4130
Qy	2514	CGCTGCTGAAGCAAAATCTTCAGCAGGCAATCTCTGATAGAGCAAGTTATTGGGAAGATT	2573
Db	4131	CAAGCAGGAGGAGAAATGTGGACCTTGAC-----TACTGGGAGAAAGCT	4172
Qy	2574	GTATAAAGATAAATTTGAGCTGCACCAGGCTGAGGAGCTTAATGCTCTTGGAAAAAGGAA	2633
Db	4173	GCTGAGGCATCACTATGAGCAACAGCAGGAAGACCTAGCCCGGAATCTAGGCAAGGGCAA	4232
Qy	2634	GAGAAGTCGCAAGCAGTTGGTATCCATTGAAGAAAGATGATCTTGCTGGTTTGGAAAGATG	2693
Db	4233	GCGGGTTCGCAAGCAAGTTAACTACAATGATGCTGCTCAGGAAGACCAAGACAACCCAGTC	4292
Qy	2694	GAGCTCTGATGGAGATGAAAGTTATGAAGCTGAGTCAACAGATGTTGAAGCAGCAGGACA	2753
Db	4293	AGAGTACTCGGTGGGTTTCAGAGGAGGAGGATGAAGACTTCGATGAACGTCCTGAAGGGCG	4352
Qy	2754	AGGAGTTCAGACGGGTCGACGGCCGTACAGAAGAAAGGGTCGGGATAAATTTGGAACCAAC	2813
Db	4353	TAGACAGTCAAAAGAGGCAGCTCCGGAA-----TGAGAAAGATAAGCCACTGCC	4400
Qy	2814	TCCGTTGATGGAAGGTGAGGGGAGATCTTTTCAGAGTACTGGGTTTCAACCAGAGTCAAAG	2873
Db	4401	TCCACTGCTGGCCCGAGTCGGGGGCAACATTGAGGTGCTGGGCTTCAACACCCGTCAGCG	4460
Qy	2874	GGCCATTTTGTACAGACTTTGATGAGGTATGG-----AGCTGGCAATTTTGTATTG	2924
Db	4461	GAAGGCTTTCCTCAATGCTGTGATGCGCTGGGGGATGCCACCACAGGATGCCTTTCACCAC	4520
Qy	2925	GAAGGAGTTTGTTCCTCGCTTAAAGCAGAAAGACCTTTGAAGAAATAAATGAATATGGAAT	2984
Db	4521	ACAGTGGCTGCTGCGGGACCTGAGGGGCAAGACTGAGAAGGAGTTTAAGGCCCTATGTGC	4580
Qy	2985	ACTCTTCTGAAGCACATTGCTGAAGAAATAGACGAGAATCTCCAACCTTTTCAGATGG	3044
Db	4581	TTTGTTCATGCGCCATCTGTGTGAGCCCTGGGGCAGACGGCTCTGAAACCTTTGCCGATGG	4640
Qy	3045	TGTGCCCCAAGGAGACTTAGAATAGAGATGTTCTAGTCAGAATTGCTCTTCTGATACT	3104
Db	4641	GGTCCCTCGGAGGGGACTGAGTCGCCACAGGTGTTGACCCGCAATGGAGTCATGTCTCT	4700
Qy	3105	AGTTCAGGAGAAGGTGAA	3122
Db	4701	CGTCAAAAAAGAGGTGCA	4718



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: September 18, 2004, 11:18:36 ; Search time 1474 Seconds  
(without alignments)  
12038.483 Million cell updates/sec

Title: US-10-049-137-1  
Perfect score: 4177  
Sequence: 1 atgagtagtttggtggagag.....caactggtaaatcaagattc 4177

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002s:\*  
7: Geneseqn2003as:\*  
8: Geneseqn2003bs:\*  
9: Geneseqn2003cs:\*  
10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4177	100.0	4177	4	AAD02817 Arabidops
2	850.8	20.4	3204	6	ABZ14721 Arabidops
3	424.6	10.2	6327	2	AAT32301 Dermatomy
4	421.4	10.1	5739	6	ABK70271 Human lun
5	421.4	10.1	6417	6	ABL64410 Stomach c
6	421.4	10.1	6417	6	ABX04169 Human mRN
7	421.4	10.1	6417	6	ABK84696 Human CDN
8	421.4	10.1	6417	7	AAD54631 Human chr
9	421.4	10.1	6417	7	AAD54632 Human chr
10	421.4	10.1	6475	4	AAI58444 Human pol
11	421.4	10.1	6475	8	ADB48415 Novel hum
12	420.8	10.1	6331	5	ABV23250 Human pro
13	420.8	10.1	6331	7	AAD54629 Human chr
14	420.8	10.1	6331	7	ACD13374 Human DNA
15	409.6	9.8	7805	7	ABX63370 Human CDN
16	387	9.3	3716	9	ADB62322 Human cod
17	387	9.3	4653	6	ABQ99469 Human cod
18	387	9.3	7849	7	AAD54630 Human chr
19	352.6	8.4	6553	4	AAI60230 Human pol
20	324.8	7.8	2673	4	ABL29909 Drosophil
C 21	324.8	7.8	4705	4	ABL29908 Drosophil
C 22	311.6	7.5	7303	4	ABL29648 Drosophil
23	296.6	7.1	6608	2	AAT42751 Chicken C

24	294.4	7.0	5947	7	AAD54627 Human chr
25	292.8	7.0	5719	7	AAD54626 Human chr
26	285.4	6.8	6364	7	ACA04019 cDNA down
27	285.4	6.8	7764	7	AAD54628 Human chr
28	279.2	6.7	6240	4	ABL06443 Drosophil
29	279.2	6.7	9933	4	ABL06442 Drosophil
30	275.8	6.6	8112	4	ABL06515 Drosophil
31	275.8	6.6	10274	4	ABL06514 Drosophil
32	264	6.3	1874	6	ABQ75913 CHD activ
33	259.6	6.2	3120	6	AAL48805 Murine an
34	259.4	6.2	11337	8	ACF36116 Human chr
35	252.6	6.0	3192	6	ABZ14466 Arabidops
36	249.8	6.0	2195	4	AAL15817 Human cDN
37	249.8	6.0	2195	6	ABL88441 Pain regu
38	249.8	6.0	6222	7	ABX70887 Novel hum
39	245.4	5.9	3174	6	ABK65338 Arabidops
40	245.4	5.9	3174	9	ADD31010 Plant vie
41	244.6	5.9	5879	7	AAD54640 Human chr
42	243	5.8	5901	7	AAD54021 Human col
43	243	5.8	5901	7	AAD54637 Human chr
44	243	5.8	9110	7	AAD54638 Human chr
45	243	5.8	9135	7	AAD54641 Human chr

ALIGNMENTS

RESULT 1  
AAD02817  
ID AAD02817 standard; cDNA; 4177 BP.  
XX  
AC AAD02817;  
XX  
DT 31-MAY-2001 (first entry)  
XX  
DE Arabidopsis thaliana PKL cDNA.  
XX  
KW PKL; chromo domain; helicase domain; DNA binding domain; CHD3 protein;  
KW zinc finger domain; transgenic plant; chromatin remodelling factor; LEC1;  
KW pickle root; developmental identity; embryo development; ss.  
XX  
OS Arabidopsis thaliana.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..4155  
FT /\*tag= a  
FT /product= "Arabidopsis thaliana PKL protein"  
FT misc\_feature 145..288  
FT /\*tag= b  
FT /note= "Encodes zinc finger domain; this region is specifically referred in claim 5"  
FT misc\_feature 343..453  
FT /\*tag= c  
FT /note= "Encodes chromo domain -I; this region is specifically referred in claim 4"  
FT misc\_feature 571..681  
FT /\*tag= d  
FT /note= "Encodes chromo domain -II; this region is specifically referred in claim 6"  
FT misc\_feature 877..2217  
FT /\*tag= e  
FT /note= "Encodes helicase domain; this region is specifically referred in claim 4"  
FT misc\_feature 3205..3285  
FT /\*tag= f  
FT /note= "Encodes DNA binding domain; this region is specifically referred in claim 4"  
XX  
PN WO200114519-A2.  
XX  
PD 01-MAR-2001.  
XX  
PF 18-AUG-2000; 2000WO-US022725.

XX 20-AUG-1999; 99US-0149975P.  
PR (PURD ) PURDUE RES FOUND.  
XX  
PI Ogas JP, Somerville CR;  
XX  
DR WPI; 2001-218433/22.  
XX P-PSDB; AAY72755.  
PT  
PT Regulating plant developmental identity comprises transforming a host  
PT cell with a nucleic acid encoding a PKL protein having at least one  
PT chromo domain, a helicase domain and a DNA binding domain.  
XX  
PS Claim 55; Page 67-76; 87pp; English.  
XX  
CC The invention relates to a method of regulating plant developmental  
CC identity which involves transforming a host cell with a nucleic acid  
CC encoding PKL (pickle) protein having atleast one chromo domain, a  
CC helicase domain, a DNA binding domain and a zinc finger domain. This  
CC method is useful for transforming a host cell with PKL proteins for  
CC regulating developmental identity, such as for regulating the transition  
CC from embryonic to post-embryonic development. The PKL proteins act as  
CC chromatin remodelling factors to repress transcription of LEC1, a protein  
CC that plays a role in regulating embryo development. The method is also  
CC used for generating transgenic plants expressing PKL proteins. The  
CC present sequence is a cDNA coding for Arabidopsis thaliana PKL protein  
CC found in pickle root. PKL functions in regulating developmental identity  
CC in host cells. It is classified as CHD3 protein based on the presence of  
CC chromo domain, SNF2-related helicase/ATPase domain, DNA-binding domain  
CC and PHD zinc finger domain  
XX  
SQ Sequence 4177 BP; 1353 A; 729 C; 1018 G; 1077 T; 0 U; 0 Other;

Query Match 100.0%; Score 4177; DB 4; Length 4177;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 4177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGTAGTTTGGTGGAGAGGCTTCGCATACGATCTGATAGGAAACCAAGTTTATAACCTA 60  
Db 1 ATGAGTAGTTTGGTGGAGAGGCTTCGCATACGATCTGATAGGAAACCAAGTTTATAACCTA 60

QY 61 GATGATCTGTGATGATGACGACTTCGTTCCCTAAAAAAGATCGAACCTTTTGAGCAAGTCGAG 120  
Db 61 GATGATCTGTGATGATGACGACTTCGTTCCCTAAAAAAGATCGAACCTTTTGAGCAAGTCGAG 120

QY 121 GCTATTGTGAGAACTGATCGGAAAGAAAATGCATGTCAGGCTTGTGGGAAAAGTACTAAT 180  
Db 121 GCTATTGTGAGAACTGATCGGAAAGAAAATGCATGTCAGGCTTGTGGGAAAAGTACTAAT 180

QY 181 CTTGTAAGCTGCAATACATGCACCTTATCGTTCCTCCATGCTAAATGCTTAGTCCACCTCTT 240  
Db 181 CTTGTAAGCTGCAATACATGCACCTTATGCGTTCCTCCATGCTAAATGCTTAGTCCACCTCTT 240

QY 241 AAAGATGCTTCGGTGGAAAATTTGGAGATGCGCTGAAATGTGTTAGTCCCTTAAACGAGATA 300  
Db 241 AAAGATGCTTCGGTGGAAAATTTGGAGATGCGCTGAAATGTGTTAGTCCCTTAAACGAGATA 300

QY 301 GATAAGATATTGGATTGTGAAATCGCTCCACAAAATCTAGTGAACAAGGTTTCCTCCGAT 360  
Db 301 GATAAGATATTGGATTGTGAAATCGCTCCACAAAATCTAGTGAACAAGGTTTCCTCCGAT 360

QY 361 GCGGAACCGAAGCAATTTTGTGAAACAGTATCTCGTGAAGTGGAGGATATATCATAC 420  
Db 361 GCGGAACCGAAGCAATTTTGTGAAACAGTATCTCGTGAAGTGGAGGATATATCATAC 420

QY 421 CTTCACTGCTCTGGGTGCCTGAGAAGGATTCAGAGGCTTATAAGTCAAAATCATCGT 480  
Db 421 CTTCACTGCTCTGGGTGCCTGAGAAGGATTCAGAGGCTTATAAGTCAAAATCATCGT 480

QY 481 TTAAAAACCAGAGTGAAACAATTTTCCACCGTCAAATGGAGTCATTCAATAACAGCGAAGAT 540  
Db 481 TTAAAAACCAGAGTGAAACAATTTTCCACCGTCAAATGGAGTCATTCAATAACAGCGAAGAT 540

QY 541 GATTTTGTGCCATACGTCCTGAGTGGACCACCTGTTGATCGGATCTTTCCTGCAGAGAG 600  
Db 541 GATTTTGTGCCATACGTCCTGAGTGGACCACCTGTTGATCGGATCTTTCCTGCAGAGAG 600

QY 601 GAAGATGGGGAGCTGGAATATCTTGTCAAATATAAAAGAGCTATCTCTATGATGAATGTTAT 660  
Db 601 GAAGATGGGGAGCTGGAATATCTTGTCAAATATAAAAGAGCTATCTCTATGATGAATGTTAT 660

QY 661 TGGGAGTCAGAAATCAGACATCTCAACCTTCCAGAATGAAATTCAAAGGTTCAAGGATGTA 720  
Db 661 TGGGAGTCAGAAATCAGACATCTCAACCTTCCAGAATGAAATTCAAAGGTTCAAGGATGTA 720

QY 721 AATCTAGAACTCGCAGAAAGTAAAGATGTTGACCATAAAAGAAATCCAGAGACTTTCAA 780  
Db 721 AATCTAGAACTCGCAGAAAGTAAAGATGTTGACCATAAAAGAAATCCAGAGACTTTCAA 780

QY 781 CAGTTTGATCATACTCCTGAAATTCCTCAAAAGGCTTGTACATCCATACCAGCTTGAGGGA 840  
Db 781 CAGTTTGATCATACTCCTGAAATTCCTCAAAAGGCTTGTACATCCATACCAGCTTGAGGGA 840

QY 841 CTTAATTTTTCGGGTTCTCGTGTCAAAAACAGACGCATGTAATCCTTGTGATGAAATG 900  
Db 841 CTTAATTTTTCGGGTTCTCGTGTCAAAAACAGACGCATGTAATCCTTGTGATGAAATG 900

QY 901 GGACTAGGCAAGACAATTCAAAGCATTGCCCTTTTAGCTTCACTTTTGGAGGAAACCTC 960  
Db 901 GGACTAGGCAAGACAATTCAAAGCATTGCCCTTTTAGCTTCACTTTTGGAGGAAACCTC 960

QY 961 ATTCCGATTTGGTAATGCTCCTCTATCGACTCTGCGTAACCTGGGAGAGAGAGTTTGCC 1020  
Db 961 ATTCCGATTTGGTAATGCTCCTCTATCGACTCTGCGTAACCTGGGAGAGAGAGTTTGCC 1020

QY 1021 ACATGGGCCCCACAGATGAACGTGGTTATGTATTTTGGCACTGCGCAAGCTCGAGCAGTT 1080  
Db 1021 ACATGGGCCCCACAGATGAACGTGGTTATGTATTTTGGCACTGCGCAAGCTCGAGCAGTT 1080

QY 1081 ATCAGAGAAACATGAGTTTACTTTATCGAAAAGATCAAAAAAGATCAAGAAAAAGAAATCT 1140  
Db 1081 ATCAGAGAAACATGAGTTTACTTTATCGAAAAGATCAAAAAAGATCAAGAAAAAGAAATCT 1140

QY 1141 GGACAATAAGTAGCGAAAGCAAGCAAAAAGAAATCAAGTTTGTATGTCTCTCACATCG 1200  
Db 1141 GGACAATAAGTAGCGAAAGCAAGCAAAAAGAAATCAAGTTTGTATGTCTCTCACATCG 1200

QY 1201 TATGATGATCAACCTAGATTACAGCAGTTCTAAAACCAATTAAGTGGGAGTGCATGATT 1260  
Db 1201 TATGATGATCAACCTAGATTACAGCAGTTCTAAAACCAATTAAGTGGGAGTGCATGATT 1260

QY 1261 GTTGATGAAGGTTCATCGACTGAAAATAAGGATTCAAAGCTGTCTCTTCAATTGACACAG 1320  
Db 1261 GTTGATGAAGGTTCATCGACTGAAAATAAGGATTCAAAGCTGTCTCTTCAATTGACACAG 1320

QY 1321 TATTCAAGTAAACCAACCGTATCTTCTGACAGGAACACCACCTTCAGAACAACTTGGATGAA 1380  
Db 1321 TATTCAAGTAAACCAACCGTATCTTCTGACAGGAACACCACCTTCAGAACAACTTGGATGAA 1380

QY 1381 CTTTTTCATGCTCATGCAATTTTCTTGATCGGGGAAGTTTGAAGTTTGGAGGAGTTCAG 1440  
Db 1381 CTTTTTCATGCTCATGCAATTTTCTTGATCGGGGAAGTTTGAAGTTTGGAGGAGTTCAG 1440

QY 1441 GAGGAGTTCAAAGATATTAATCAAGAGGAGCAGATCTCAAGGTTTGACAAAAATGTTGGCT 1500  
Db 1441 GAGGAGTTCAAAGATATTAATCAAGAGGAGCAGATCTCAAGGTTTGACAAAAATGTTGGCT 1500

QY 1501 CCACATTTGCTCAGAAAGGTAAAAAAGACGTAAATGAAAGACATGCCCCCCAAAAAGGAG 1560  
Db 1501 CCACATTTGCTCAGAAAGGTAAAAAAGACGTAAATGAAAGACATGCCCCCCAAAAAGGAG 1560

QY 1561 CTCATTTTGGGTGTTGATCTGAGCAGTCTGCAGAAAAGATATTACAAAGCTATTTTACC 1620  
Db 1561 CTCATTTTGGGTGTTGATCTGAGCAGTCTGCAGAAAAGATATTACAAAGCTATTTTACC 1620

QY 1621 CGTAATTATCAAGTATTGACAAAAAGGGAGGTGCTCAAATTTCCCTTAATAACATTATG 1680  
Db 1621 CGTAATTATCAAGTATTGACAAAAAGGGAGGTGCTCAAATTTCCCTTAATAACATTATG 1680  
QY 1681 ATGGAATTACGAAAAGTATGCTGCCATCCTTATATGCTAGAGGGTGTGAGCCAGTTATT 1740  
Db 1681 ATGGAATTACGAAAAGTATGCTGCCATCCTTATATGCTAGAGGGTGTGAGCCAGTTATT 1740  
QY 1741 CACGACGCAAAATGAAGCTTTCAAAACAACTTTTGGAGTCTTGTGGAAGCTGCAACITCTA 1800  
Db 1741 CACGACGCAAAATGAAGCTTTCAAAACAACTTTTGGAGTCTTGTGGAAGCTGCAACITCTA 1800  
QY 1801 GATAAAATGATGGTCAAACTGAAAGAGCAAGGACACAGAGTCCTAAATATACACACAGTTT 1860  
Db 1801 GATAAAATGATGGTCAAACTGAAAGAGCAAGGACACAGAGTCCTAAATATACACACAGTTT 1860  
QY 1861 CAGCATATGCTGGACTTACTTTGAAGACTACTGTACCCATAGAAATGGCAGTACGAGCGA 1920  
Db 1861 CAGCATATGCTGGACTTACTTTGAAGACTACTGTACCCATAGAAATGGCAGTACGAGCGA 1920  
QY 1921 ATTGATGAAAAGGTTGGCGAGCTGAGCGGCAAAATACGCATAGATCGGTTCAATGCCAAA 1980  
Db 1921 ATTGATGAAAAGGTTGGCGAGCTGAGCGGCAAAATACGCATAGATCGGTTCAATGCCAAA 1980  
QY 1981 AATTCTAAACAAGTTTTGTTTTTTTGTCTCTCCACAAGAGCTGGTGGCTTAGGAATAAACTT 2040  
Db 1981 AATTCTAAACAAGTTTTGTTTTTTTGTCTCTCCACAAGAGCTGGTGGCTTAGGAATAAACTT 2040  
QY 2041 GCAACGGCTGATACAGTAATCATTTATGACAGTGACTGGAATCCTCATGCTGATCTCAA 2100  
Db 2041 GCAACGGCTGATACAGTAATCATTTATGACAGTGACTGGAATCCTCATGCTGATCTCAA 2100  
QY 2101 GCAATGGCTAGAGCTCATCGACTTGGCCAAAACAATAAGGTGATGATTTATAGGCTCATA 2160  
Db 2101 GCAATGGCTAGAGCTCATCGACTTGGCCAAAACAATAAGGTGATGATTTATAGGCTCATA 2160  
QY 2161 AACCAGGCACCATTTGAAGAAAGGATGATGCAATTGACTAAAAAGAAAATGGTTCTAGAG 2220  
Db 2161 AACCAGGCACCATTTGAAGAAAGGATGATGCAATTGACTAAAAAGAAAATGGTTCTAGAG 2220  
QY 2221 CATCTTGTGTTGGGAACTCAAAACACAAAACATTAATCAGGAAGAGTTAGATGACATC 2280  
Db 2221 CATCTTGTGTTGGGAACTCAAAACACAAAACATTAATCAGGAAGAGTTAGATGACATC 2280  
QY 2281 ATCAGGTATGGATCAAAGGAGCTTTTGCTAGTGAAGATGATGAAGCAGGAAAGTCTGGA 2340  
Db 2281 ATCAGGTATGGATCAAAGGAGCTTTTGCTAGTGAAGATGATGAAGCAGGAAAGTCTGGA 2340  
QY 2341 AAAATTCAATATGATGATGCGGCTATAGACAAAATGCTTGATCGTGATCTCGTGGAGGCA 2400  
Db 2341 AAAATTCAATATGATGATGCGGCTATAGACAAAATGCTTGATCGTGATCTCGTGGAGGCA 2400  
QY 2401 GAGGAAGTCTCAGTGGATGATGAAGAGGAGAATGGATTTTAAAGGCTTTCAAGGTGGCT 2460  
Db 2401 GAGGAAGTCTCAGTGGATGATGAAGAGGAGAATGGATTTCTTAAAGGCTTTCAAGGTGGCT 2460  
QY 2461 AATTTTGAATATATAGATGAAAATGAGGCAGCAGCATTTAGAGGCACAGAGTCGCTGCT 2520  
Db 2461 AATTTTGAATATATAGATGAAAATGAGGCAGCAGCATTTAGAGGCACAGAGTCGCTGCT 2520  
QY 2521 GAAAGCAAACTTTCAGCAGGCAATCTGTATAGAGCAAGTATTTGGGAAGAGTTGTTAAAA 2580  
Db 2521 GAAAGCAAACTTTCAGCAGGCAATCTGTATAGAGCAAGTATTTGGGAAGAGTTGTTAAAA 2580  
QY 2581 GATAAAATTTGAGCTGCACCAGGCTGAGGAGCTTAATGCTCTTGGAAAAAGGAGAGT 2640  
Db 2581 GATAAAATTTGAGCTGCACCAGGCTGAGGAGCTTAATGCTCTTGGAAAAAGGAGAGT 2640  
QY 2641 CGCAAGCAGTTGGTATCCATTGAAGAAGATGATCTTGCTGTTTGGAAAGATGTGAGCTCT 2700  
Db 2641 CGCAAGCAGTTGGTATCCATTGAAGAAGATGATCTTGCTGTTTGGAAAGATGTGAGCTCT 2700  
QY 2701 GATGGAGATGAAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGACAAGGAGTT 2760

Db 2701 GATGGAGATGAAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGACAAGGAGTT 2760  
QY 2761 CAGACGGGTTCGACGGCCGTACAGAAAGAAAGGTCGCGATAAATTTGGAAACCAACTCCGTTG 2820  
Db 2761 CAGACGGGTTCGACGGCCGTACAGAAAGAAAGGTCGCGATAAATTTGGAAACCAACTCCGTTG 2820  
QY 2821 ATGGAAGGTGAGGGGAGATCTTTTCAGAGTACTGGGTTTCAACACAGAGTCAAAAGGGCCATT 2880  
Db 2821 ATGGAAGGTGAGGGGAGATCTTTTCAGAGTACTGGGTTTCAACACAGAGTCAAAAGGGCCATT 2880  
QY 2881 TTTGTACAGACTTTGATGAGGTATGGAGCTGGCAATTTTGATTTGGAAGGAGTTTGTTCCT 2940  
Db 2881 TTTGTACAGACTTTGATGAGGTATGGAGCTGGCAATTTTGATTTGGAAGGAGTTTGTTCCT 2940  
QY 2941 CGCTTAAAGCAGAAGACCTTTGAAGAAAATAAATGAATATGGAATACTCTTCTTGAAGCAC 3000  
Db 2941 CGCTTAAAGCAGAAGACCTTTGAAGAAAATAAATGAATATGGAATACTCTTCTTGAAGCAC 3000  
QY 3001 ATTGCTGAAGAAAATAGACGAGAAATTCCTCAACCTTTTCAGATGGTGTGCCCCAAGGAAGGA 3060  
Db 3001 ATTGCTGAAGAAAATAGACGAGAAATTCCTCAACCTTTTCAGATGGTGTGCCCCAAGGAAGGA 3060  
QY 3061 CTTAGAAATAGAAGATGTTCTAGTCAGAAATTCCTTCTTGATACTAGTTTCAGGAGAAGGTG 3120  
Db 3061 CTTAGAAATAGAAGATGTTCTAGTCAGAAATTCCTTCTTGATACTAGTTTCAGGAGAAGGTG 3120  
QY 3121 AAATTTGTAGAAGATCATCCAGGGAAAACCTGTTTTCCCTCTCGCATTTCTTGAAGATTC 3180  
Db 3121 AAATTTGTAGAAGATCATCCAGGGAAAACCTGTTTTCCCTCTCGCATTTCTTGAAGATTC 3180  
QY 3181 CCCGGACTGAGAAAGTGGRAAAAATTTGGAAGGAGGAACATGACAAGATAATGATACGTGCT 3240  
Db 3181 CCCGGACTGAGAAAGTGGRAAAAATTTGGAAGGAGGAACATGACAAGATAATGATACGTGCT 3240  
QY 3241 GTTTTAAAGCATGGGTACCGACGGTGGCAAGCTATTGTTGATGACAAAGAGTTGGGGATC 3300  
Db 3241 GTTTTAAAGCATGGGTACCGACGGTGGCAAGCTATTGTTGATGACAAAGAGTTGGGGATC 3300  
QY 3301 CAAGAGCTTATCTGCAAGAAATGAAATTTCCCTCACATAAGTTTGTCTGCTGAACAA 3360  
Db 3301 CAAGAGCTTATCTGCAAGAAATGAAATTTCCCTCACATAAGTTTGTCTGCTGAACAA 3360  
QY 3361 GCTGGTTTGCAGGGGCAGAAATGGTAGTGGGGCTCTAATCCGGGAGCACAGACTAACCCAG 3420  
Db 3361 GCTGGTTTGCAGGGGCAGAAATGGTAGTGGGGCTCTAATCCGGGAGCACAGACTAACCCAG 3420  
QY 3421 AATCCTGGAAGCGTTATTACTTGGGAAACAATAATGCTTCTGCTGATGGGCTCAAGTAAAC 3480  
Db 3421 AATCCTGGAAGCGTTATTACTTGGGAAACAATAATGCTTCTGCTGATGGGCTCAAGTAAAC 3480  
QY 3481 TCGATGTTCTATTATCGGGACATGCAGAGACGACTTGTGAGTTTGTGAAAAAGCGAGTT 3540  
Db 3481 TCGATGTTCTATTATCGGGACATGCAGAGACGACTTGTGAGTTTGTGAAAAAGCGAGTT 3540  
QY 3541 CTGCTTTTGGAGAAGGCGATGAATTTATGAATACGAGAGGAAATATTATGGACTTGGTGGC 3600  
Db 3541 CTGCTTTTGGAGAAGGCGATGAATTTATGAATACGAGAGGAAATATTATGGACTTGGTGGC 3600  
QY 3601 TCATCATCTATCCCTACTGAAGAAACAGAAAGCTGAACCAAGATCGCTGACACAGTGGGA 3660  
Db 3601 TCATCATCTATCCCTACTGAAGAAACAGAAAGCTGAACCAAGATCGCTGACACAGTGGGA 3660  
QY 3661 GTGAGCTTTATTGAGGTTGATGATGAAATGCTTTGATGGACTTCTTAAGACTGATCCTATC 3720  
Db 3661 GTGAGCTTTATTGAGGTTGATGATGAAATGCTTTGATGGACTTCTTAAGACTGATCCTATC 3720  
QY 3721 ACTTCAGAAAGAAATTTATGGGGCTGCTGTTGACAACCAACGCGGGTCGAAATAGCT 3780  
Db 3721 ACTTCAGAAAGAAATTTATGGGGCTGCTGTTGACAACCAACGCGGGTCGAAATAGCT 3780  
QY 3781 CAACATTATAACCCAGATGTGCAAACTTCTTGATGAGAACGCTCGGGAATCAGTCCAAAGCA 3840



Db 3781 CAACATTATAACACAGATGTGCAAACTTCTTGATGAGAACGCTCGGGAATCAGTCCAAGCA 3840  
QY 3841 TATGTAACAACAACCAACCCGAGTACCAAGGTGAATGAGAGCTTCCGTGCACCTCAAATCT 3900  
Db 3841 TATGTAACAACAACCAACCCGAGTACCAAGGTGAATGAGAGCTTCCGTGCACCTCAAATCT 3900  
QY 3901 ATCAATGTAACATTAACACAAATCCTTTCGATTACATCTGATCAATCCAAGTCACATGAA 3960  
Db 3901 ATCAATGTAACATTAACACAAATCCTTTCGATTACATCTGATCAATCCAAGTCACATGAA 3960  
QY 3961 GACGACACCAAGCCAGACCTAAACAATGTTGAGATGAAGGACACGGCCGAAGAAACAAA 4020  
Db 3961 GACGACACCAAGCCAGACCTAAACAATGTTGAGATGAAGGACACGGCCGAAGAAACAAA 4020  
QY 4021 CCGTTAAGAGGTGGCGTCGTCGATCTGAATGTGGTGAGGAGGAGAACATTGCTGAA 4080  
Db 4021 CCGTTAAGAGGTGGCGTCGTCGATCTGAATGTGGTGAGGAGGAGAACATTGCTGAA 4080  
QY 4081 GCTAGTGAAGTGTGATGTAAAAATGGAAGAACCCAAAGAGAGAGAACCCAAAGAAC 4140  
Db 4081 GCTAGTGAAGTGTGATGTAAAAATGGAAGAACCCAAAGAGAGAGAACCCAAAGAAC 4140  
QY 4141 ATGGTCGTTGATTGACTCAACTGGTAAATCAAGATT 4177  
Db 4141 ATGGTCGTTGATTGACTCAACTGGTAAATCAAGATT 4177

RESULT 2  
ABZ14721  
ID ABZ14721 standard; DNA; 3204 BP.  
XX  
AC ABZ14721;  
XX  
DT 21-JAN-2003 (first entry)  
XX  
DE Arabidopsis thaliana stress regulated gene SEQ ID NO 2526.  
XX  
KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.  
XX  
OS Arabidopsis thaliana.  
XX  
PN WO200216655-A2.  
XX  
PD 28-FEB-2002.  
XX  
PF 24-AUG-2001; 2001WO-US026685.  
XX  
PR 24-AUG-2000; 2000US-0227866P.  
PR 26-JAN-2001; 2001US-0264647P.  
PR 22-JUN-2001; 2001US-0300111P.  
XX  
PA (SCRI ) SCRIPPS RES INST.  
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX  
PI Harper JF, Kreps J, Wang X, Zhu T;  
XX  
DR WPI; 2002-304127/34.

XX  
PT Identifying a stress condition to which a plant cell has been exposed and  
PT producing plants with increased tolerance to these abiotic stresses.  
XX  
PS Claim 144; SEQ ID NO 2526; 577pp + Sequence Listing; English.  
XX  
CC The invention relates to identifying a stress condition to which a plant  
CC cell has been exposed, comprising: (a) contacting nucleic acid  
CC representative of expressed polynucleotides in the plant cell with an  
CC array or probes representative of the plant cell genome; and (b)  
CC detecting a profile of expressed polynucleotides in the plant cell  
CC characteristic of a stress response. The method is useful in the  
CC production of transgenic plants, cells and seeds and in producing plants  
CC with increased tolerance to abiotic stress. The present sequence is that  
CC of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used  
CC in methods of the invention. Note: The sequence data for this patent is

CC not represented in the printed specification but is based on sequence  
CC information supplied to Derwent by the European Patent Office  
XX  
SQ Sequence 3204 BP; 1048 A; 561 C; 784 G; 811 T; 0 U; 0 Other;  
Query Match 20.4%; Score 850.8; DB 6; Length 3204;  
Best Local Similarity 61.0%; Pred. No. 8.2e-217;  
Matches 1892; Conservative 0; Mismatches 697; Indels 513; Gaps 11;  
QY 280 GTTAGTCCTCTTAACGAGATAGATAGATATTGGATTGTGAAATGCGTCTACAAAATCT 339  
Db 121 GATTGCTCTTTGGCGAGATTGAAAAGATTTTGGACCGGTGAATGGCGTCTACTGCACT 180  
QY 340 AGTGAACAAGGTTCTCTCCGAT--GCGGAACCGAAGCCAAATTTTGTGAAACAGTATCTC 396  
Db 181 AACAAATCCGAATTTCTTGACAAACGGGACACCAACCCCTAGTTGTGTAACAATATCTT 240  
QY 397 GTGAAGTGAAGGATTATCATACCTTCACCTGCTCTTGGTGCTGAGAAGGAGTCCAG 456  
Db 241 GTGAAGTGAAGGTTATCGTACTTGCACCTGCTCTTGGTGCTGAGAAGGAGTCCAG 300  
QY 457 AAGGCTTATAAGTCAAAATCATCGTT-----TAAAAACCCAGAGTGAACAATTTTCACCGT 510  
Db 301 AAGGCTTACAAGTCCCATCCCTCAATTTAAAACTAAATTAAGGTTGACCAAGTTTAAACGCT 360  
QY 511 CAAATGGAGTCATTCA-----ATAACAGCGAAGATGATTTTGTGCCATACGTCCTGAG 564  
Db 361 GCTATGGACGTATTTCATCGCTGAGAACGGAGCTCATGAGTTATCGCCATTCTGCTCGAA 420  
QY 565 TGGACCACTGTTGATCGGATTCTTGCCTGCAGAGAGGAAGATGGGAGCTGGAATATCTT 624  
Db 421 TGGAAAACTGTTGACCCGATTATTGCTTCCCGGAAGGAGATGATGGGAGGAGTATCTG 480  
QY 625 GTCAAAATATAAAGAGCTATCCTATGATGAATGTTATTGGAGTCAGAATCAGACATCTCA 684  
Db 481 GTGAAGTACAAAGAACTTTCTTACAGAAATAGTTACTGGAATCTGAATCTGACATCTCA 540  
QY 685 ACCTTCCAGAAATGAAATTCAAAGGTTCAAGGATGTAATTTCTAGAACTCGCAGAAATAA 744  
Db 541 GATTTCAGAAATGAAATCCAAAGGTTAAGGATATAAACTTAGCTCTCGTAGAGATAA 600  
QY 745 GATGTTGACCATAAAAGAAATCCAGAGACTTTTCAACAGCTTTGATCATACTCCTGAATTC 804  
Db 601 TATGTCGAAAATGAGAGAAATCGAGAGGAATTCAAACAGTTTGAATCTTACTCTGAATTT 660  
QY 805 CTCAAAGGCTTGTACATCCATACAGCTTGAGGGAATTAATTTTTCGGTCTCGTGG 864  
Db 661 CTTACAGGCACATTGCATACATACAGCTCGAAGGCTGAATTCCTAAGGTATTCATGG 720  
QY 865 TCAAAACAGACGCATGTAATCCTTGTCTGATGAAATGGGACTAGGCAAGACAATTCAAAGC 924  
Db 721 TCCAAAGAAAACCAATGTAATCCTTGGGATGAAATGGGACTAGGCAAAACTATTTCAGAGC 780  
QY 925 ATTGCCCTTTTAGCTTCACCTTTTGGAGGAAACCTCATTCGCAATTTGGTAATTTGCTCCT 984  
Db 781 ATTGCCCTTTTGGCCTCTTTGTTCCGAGGAAACTTATCTCCTCATCTAGTTGTTGCTCCT 840  
QY 985 CTATCGACTCTGCGTAACCTGGGAGAGAGAGTTTGGCCACATGGGCCCCCAGATGAACGTG 1044  
Db 841 CTCTCAACCATCCGTAACCTGGGAGAGGGAATTTGCCACCTGGGACCACATATGAATGTG 900  
QY 1045 GTTATGTATTTTGGCACTGGCAAGCTCGAGCAGTTATCAGAGAACATGAGTTTACTTTA 1104  
Db 901 GTTATGTACACTGGGGAATTCGGAAGCAGCTGATGTTATATGGAACATGAGTTCTACTTC 960  
QY 1105 TCGAAAGATCAAAAAAAGATCAAGAAAAAGAAATCTGGACAAATAAGTAGCGAAAGCAAG 1164  
Db 961 TCTGAAGGGCGTAAAAGGTTAAAG-----984  
QY 1165 CAAAAAAGAAATCAAGTTTGTATGTCCTCTCACATCGTATGAGATGATCAACCTAGATTCA 1224  
Db 985 -----984

Qy 1225 GCAGTCTCTAAACCAATTAAAGTGGAGTGCAATGATTGTTGATGAAGGTCATCGACTGAAA 1284  
Db 985 ----- 984  
Qy 1285 AATAAGGATTCAAAGCTGTTCTCTTCATTGACACAGTATTCAAGTAACCAACCGTATTCTT 1344  
Db 985 ----- 984  
Qy 1345 CTGACAGGAACACCACTTCAGAACAACTTGGATGAACTTTTCATGCTCATGCAATTTTCTT 1404  
Db 985 ----- 984  
Qy 1405 GATCGGGGAAGTTTGGAAGTTTGAGGAGTTTCCAGGAGGAGTTCAAAGATATATTAATCAA 1464  
Db 985 ----- 984  
Qy 1465 GAGGAGCAGATCTCAAGTTTGCACAAAATGTTGGCTCCACATTTGCTCAGAAGGGTA AAA 1524  
Db 985 -----AAA 987  
Qy 1525 AAAGACGTAATGAAAGACATGCCCCCAAAAAGGAGCTCATTTTTCGTTGATCTGAGC 1584  
Db 988 GACGTGCTCAAGGATAAGGTGCCCCCAAGAAGGAGCTCATTTTTCGTTGATATGAGC 1047  
Qy 1585 AGTCTGCAGAAAAGAATATTACAAAGCTATTTTACCCGTAATATTCAAGTATTGACAAAA 1644  
Db 1048 AGCAGCAAAAAGAAGTGTAAGAAGCTGTTATTACCAATAATTATCAAGTCTTGACAAAG 1107  
Qy 1645 AAGGAGGTGCTCAAAATTTCCCTTAATAACATTATGATGGAATTACGAAAAGTATGCTGC 1704  
Db 1108 AAACGAGATGCTAAAATATC-----TAACGTGCTAATGAAATACGCCAAGTTTGCTCT 1161  
Qy 1705 CATCCTTATATGCTAGAGGGTGTGAGCCAGTTATTACAGACGCAAAATGAAGCTTTCAAA 1764  
Db 1162 CATCCGTATCTGTTACCAGACTTTGAGCCAAAGATTGAAGACGCAAAATGAAGCTTTTACA 1221  
Qy 1765 CAACTTTGGAGTCTTGTGGAAGCTGCAACTTCTAGATAAAATGATGGTCAAACTGAAA 1824  
Db 1222 AAACATTGGAAAGCTTCCGGAAGTTGCAACTGTTAGACAAAATGATGGTCAAACTCAA 1281  
Qy 1825 GAGCAAGCACACAGATCCTTAATATACACACAGTTTTCAGCATATGCTGGACTTACTTGAA 1884  
Db 1282 GAGCAGGGCACAGAGTGCTAATCTATACCCAAATTCACACACACCCCTTACCTTTTGAA 1341  
Qy 1885 GACTACTGTACCCATAAGAAATGGCAGTACGAGCGAATTGATGGAAGGTTGGCGGAGCT 1944  
Db 1342 GACTACTTCACTTTCAAGAAATTGGAATTATGAACGGATTGATGGAAGATCAGTGGACCT 1401  
Qy 1945 GAGCGCAAAATACGCATAGATCGGTTCAATGCCAAAAATTCFAACAAAGTTTGTTTTTTG 2004  
Db 1402 GAGAGGCAAGTAAGAATAGATCGGTTTAATGCAGAAAACTCGAACCGGTTCTGTTTTTTG 1461  
Qy 2005 CTCTCCACAAAGAGCTGGTGGCTTAGGAATAAATCTTGCAACGCGTGATACAGTAATCAAT 2064  
Db 1462 CTCAGTACAAGAGCTGGTGGATAGGGATAAACCTTGCAACAGCTGATACAGTCATCAT 1521  
Qy 2065 TATGACAGTGACTGGAATCCTCATGCTGATCTTCAAGCAATGSGTAGAGCTCATCGACTT 2124  
Db 1522 TATGACAGTGATTGGAATCCTCATGCTGATCTCCAGGCTATGGCTAGAGTTTCATCGGCT 1581  
Qy 2125 GGCCAAACAAATAAGGTGATGATTATAGGCTCATAAACCGAGGCACCATTTGAAGAAAG 2184  
Db 1582 GGCCAAACAAATAAGGTAATGATCTACAGACTGATTCCAAAAGGTACTGTAGAAGAACGG 1641  
Qy 2185 ATGATGCAATTGACTAAAAAGAAAATGGTTCTAGAGCATCTTGTGTTGTTGGAAAAC 2244  
Db 1642 ATGATGGAGATAACCAAGAACAAAGATGCTTCTAGAGCATCTCGTCGTGGG-----A 1692  
Qy 2245 ACACAAAACATTAATCAGGAAGAGTTTAGATGACATCATCAGGTATGGATCAAGGAGCTT 2304  
Db 1693 AAACAACATCTTTGTCAGGACGAGTTAGATGACATCATCAAAATATGTTTCCAAGGAAC 1752  
Qy 2305 TTTGCTAGTGAAGATGATGAAGCAGGAAAAGTCTGGAAAAATTCATTATGATGATGCGGCT 2364

Db 1753 TTCTCTGAAGAGAATGATGAAGCAGGGAGGTCGGAAAAATCCATTACGATGATGCTGCT 1812  
Qy 2365 ATAGACAAAATTGCTTGATCCTGATCTCGTGGAGGCAGAGGAAGTCTCAGTGGATGATGAA 2424  
Db 1813 ATCGAACAGTTGCTAGATCGTAATCATGTAGATGCTGTGGAGGTCTCAGTGGATGATGAG 1872  
Qy 2425 GAGGAGAATGGATTCTTAAAGGCTTTCAAGGTGGCTAAATTTTGAATATATAGATGAAAA 2484  
Db 1873 GAGGAAAAC 1932  
Qy 2485 GAGGCAGCAGCATTAGAGGCACAGAGAGTCGCTGCTGTAAGAACAAAATCTTCAGCAGGCAAT 2544  
Db 1933 AATGAGGCTGCAGCATTAGAGGAGGCACAAGCTATAGAAAAACAATCTTTCAGTAAGAAAT 1992  
Qy 2545 TCTGATAGAGCAAGTTATTGGGAAGAGTTGTTTAAAGATAAATTTGAGCTGCACCAGGCT 2604  
Db 1993 GCTGATAGAACAAAGTCACTGGAAGATTTGCTGAAAAGACAAATATGAAGTGCAACAAGCT 2052  
Qy 2605 GAGGAGCTTAATGCTCTTGGAAAAAGGAAGAGAAAGTCGCAAGCAGTTGGTATCCATTGAA 2664  
Db 2053 GAGGAGCTCAGTGCTCTTGGAAAAGAGGAAGAAAACGGCAAGCAGGTGATGTA---TGCT 2109  
Qy 2665 GAAGATGATCTTGTGTTTGGAAAGATGTGAGCTCTGATGGAGATGAAAAGTTATGAAGCT 2724  
Db 2110 GAAGATGATCTTGTGTTTGAAGAAATCTCTGATGAGGAGGACGAATATTGTCTTGAT 2169  
Qy 2725 GAGTCAACA-----GATGGTGAAGCAGCAGGACAA 2754  
Db 2170 GATTTAAAAGTGACCTCTGATGAGGAGGAGGAAGCAGATGAACCCGAAGCAGCTCGACAG 2229  
Qy 2755 GGAGTTC--AGACGGGTCGACGGCCGTACAGAAAGAAAGGTCGCGATAATTTGGAACCA 2811  
Db 2230 AGAAAGCCGAGGACGGTTACAAGGCCCTTACAGAAAAAGGCTCGCGATAATTCGGAGGAA 2289  
Qy 2812 ACTCCGTTGATGGAAGGTGAGGGGAGATCTTTTCAGAGTACTGGGTTTCAACCAGAGTCAA 2871  
Db 2290 ATACCTTTGATGGAAGGCGAAGGGAGATATCTCATGCTGCTGGGATTTAACGAGACTGAG 2349  
Qy 2872 AGGGCCATTTTGTACAGACTTTGATGAGGTATGGAGCTGGCAATTTTGTGTTGGAAGGAG 2931  
Db 2350 AGGGATATATTTTACGTACGTTTAAAGAGGTATGGAATCTTATTTTGAACAACATTTGCC 2409  
Qy 2932 TTTGTTCTCGCTTAAAGCAGAAAGACCTTTTGAAGAAAATAAATGAATATGGAATACTCTTC 2991  
Db 2410 GAAAAACCTACCGATAATTTCTACAAATTTTAAAGTTATTACGGCAATGGT----- 2459  
Qy 2992 TTGAAGCACATTGCTGAAGAAATAGACGAGAATTTCTCCAACCTTTTCAGATGGTGTGCC 3051  
Db 2460 -----CTATGCAGATGGAGTTCCC 2478  
Qy 3052 AAGGAAGGACTTAGAATAGAAAGTGTCTAGTCAGAAATTGCTCTTCTGATACTAGTTTCAG 3111  
Db 2479 AAAGAAGGGATAAGCAGTGACGAGCTACTAGTGAGCATGACTTTTATGATGCTAGTAAAG 2538  
Qy 3112 GAGAAAGGTGAAAATTTGTAGAAGATCATCCAGGGAAAACCTGTTTTTCCCCCTCTCGCATCTT 3171  
Db 2539 GAGAAAGTGTCAATTTTGGACAACCATCCGACAGCACCTGTTTT---CTCTAACTACGTC 2595  
Qy 3172 GAAAAGATTCCCCGACTGAGAAAGTGGAAAAATTTGGAAGAGGAAACATGACAAGATAATG 3231  
Db 2596 ATCAGTAAGTACAATTTTGAGAAATGGAGCATTTTCTAAGGAAGAAACATGATAGGATACTG 2655  
Qy 3232 ATACGTGCTGTTTAAAGCATGGGTACGGACGGTGGCAAGCTATTGTTGATGACAAAAGAG 3291  
Db 2656 ATTCTGCTGTTTCCAAGCATGGCTATGGTAGATGGGTGGCCATCGTTGAAGACGGAAGAG 2715  
Qy 3292 TTGGGGATCCAAGAGCTTATCTGCAAGAAATTTGAATTTCCCT 3333  
Db 2716 ATAGGGTTCCAAGAGGTTGCTGCAAAAGACCTGAATATATCCCT 2757





Db 3424 TTATCTATGACTCTGACTGGAAACCCCAATAATGACATTCAGGCCTTTAGCAGAGCTCACC 3483  
Qy 2120 GACTTGGCCAAACAAATAAGGTGATGATTTATAGGCTCATAAACCGAGGCCACCATTTGAAG 2179  
Db 3484 GGATTGGGCAAAATAAAAAGGTAATGATCTACCGTTTGTACCCGTCGTCAGTGGAGG 3543  
Qy 2180 AAAGGATGATGCAATTGACTAAAGAAATAATGTTCTTAGAGCATCTTGTGTTGG----- 2234  
Db 3544 AGCGCATGACGCGAGGTGGCAAGAAAGAAATAATGATGCTGACGCATCTAGTGGTGCAGGCTG 3603  
Qy 2235 ----GAAACTCAAAACACAAACATTAATCAGGAAGAGTTAGATGACATCATCAGGTATG 2290  
Db 3604 GGCTGGGCTCCAAGACTGGATCTATGTCCAACAGGAGCTTGATGATATCTCCTCAAATTTG 3663  
Qy 2291 GATCAAAGGAGCTTTTGTGCTAGTGAAGATGATGAAGCAGGAAAGTCTGGAAAAATTCATT 2350  
Db 3664 GCACTGAGGAACCTATTCAAGGATGAAGCCACTGATGGAGGAGGAGCAACAAGAGGGAG 3723  
Qy 2351 ATGATGATGCGGCTATAGACAAATTGCTTGA-----TCGTGATCTCGTGGAGGCAGAGG 2404  
Db 3724 AAGATAGCAGTGTATTCCACTACGATGATAAGGCCATTGAACGGCTGCTAGACCCGTAACC 3783  
Qy 2405 AAGTCTCAGTGGATGATGAAGAGGAGAAATGATCTTAAAGGCTTTCAAGGTGGCTAATT 2464  
Db 3784 AGGATGAGACTGAAGACACAGAATTGCAGGCGCATGAATGAATATTTGAGCTCATTCAAAG 3843  
Qy 2465 TTGAATATATAGATGAAAAATGAGGCAGCAGCATTAGAGGCACAGAGATCGTCTGATAA 2524  
Db 3844 TGGCCCACTATGTGGTACGGGAAGAAAGAAATGGGGAGGAAGAGACGTAGAACGGGAAA 3903  
Qy 2525 GCAAATCTTCAGCAGGCAATTCTGTATAGACAGCAAGTTATTGGGAAGAGTTGTTAAAGATA 2584  
Db 3904 TCATTAAACAGGAAGAAAGTGTGGATCTTGA---CTACTGGGAGAAATGCTGCGGCACC 3960  
Qy 2585 AATTGAGCTGCACCAGGCTGAGGAGCTTAATGCTCTTGGAAAAAGGAAGAGAAAGTCGCA 2644  
Db 3961 ATTATGAGCAGCAGCAAGAAGATCTAGCCGAAATCTGGGCAAGGAAAAAGAAATCCGTA 4020  
Qy 2645 AGCAGTTGGTATCCATTGAAG---AAGATGATCTTGTGTTTGGAAAGATGTGAGCTCTG 2701  
Db 4021 AACAGGTCAACTACAATGATGGCTCCAGGAGGACCGAGATTGGCAGGACGACCATCGC 4080  
Qy 2702 ATGGAGATGAAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGACAAAGGATTC 2761  
Db 4081 ACAACCACTCGGATTACTCAGTGGCTTCAGAGGAAGGTGATGAAGACTTTGATGAACGTT 4140  
Qy 2762 AGACGGTGCAGCGCCGTACAGAAGAAAGGTGCGGATAATTTGGAAACCAA----- 2812  
Db 4141 CAGAAAGCTCCCCGTAGGCCCCAGTCGTAAAGGCCCTGCGGAATGATAAGATAAGCCATTGC 4200  
Qy 2813 CTCGGTTGATGGAAGGTGAGGGAGATCTTTTCAGAGTACTGGGTTTCAACAGAGTCAAA 2872  
Db 4201 CTCCTCTGTTGGCCCGTGTGTGGGAAATATGAAGTACTTGGTTTAAATGCTCGTCAGC 4260  
Qy 2873 GGGCCATTTTGTACAGACTTTTGATGAGGTATGG-----AGCTGGCAATTTTGATT 2923  
Db 4261 GAAAAGCCTTTCTTAATGCAATTATGCGATATGGTATGCCACCTCAGGATGCTTTTACTA 4320  
Qy 2924 GGAAGGAGTTTGTCTCGCTTAAGCAGAGACCTTTGAAGAAATAGACAAATAATGAATATGAA 2983  
Db 4321 CCCAGTGGCTTTGAAGAGACCTGCGAGGCAATATCAGAGAAGAGTTCAAGGCATATGCT 4380  
Qy 2984 TACTCTTCTTGAAGCACATTGCTGAAGAAATAGACGAGAATTTCTCCAACCTTTTCAGATG 3043  
Db 4381 CTCTTTTCATCGGCATTTATGTAGCCCGGGCGCAGATGGGGCTGAGACCTTTGCTGATG 4440  
Qy 3044 GTGTGCCCAAGGAAGGACTTAGAATAGAGATGTTCTTAGTCAGAAATGCTCTTCTGTATAC 3103  
Db 4441 GTGTCCCCCGAGAGGCGCTGTCTCGCCAGCATGTCTTACTAGAAATGCTGTATGTCTT 4500  
Qy 3104 TAGTTCAGGAGAGGTGAAATTTGTAGAGATCATCCAGGG 3144

Db 4501 TGATTGCGCAAGAAGTTTCAGGAGTTTGAACATGTTAATGGG 4541  
RESULT 4  
ABK70271  
ID ABK70271 standard; cDNA; 5739 BP.  
XX  
AC ABK70271;  
XX  
DT 15-JUL-2002 (first entry)  
XX  
DE Human lung cancer associated full length cDNA DMSM-6.  
XX  
KW Human; ss; gene; lung cancer; cytostatic; tumour; vaccine.  
XX  
OS Homo sapiens.  
XX  
PN WO200224057-A2.  
XX  
PD 28-MAR-2002.  
XX  
PF 20-SEP-2001; 2001WO-US042232.  
XX  
PR 22-SEP-2000; 2000US-0234837P.  
PR 10-OCT-2000; 2000US-0239440P.  
PR 29-JUN-2001; 2001US-0301928P.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Benson DR, Mohamath R, Lodes MJ;  
XX  
DR WPI; 2002-372001/40.  
XX  
PT New tumor lung proteins and nucleic acids encoding the proteins, useful as vaccines and for treating, preventing, diagnosing or monitoring lung cancer.  
XX  
PS Claim 1; Page 146-148; 189pp; English.  
XX  
CC The invention relates to an isolated polynucleotide comprising a sequence selected from 183 human DNA sequences (appearing as ABK70130-ABK70312), or their fragments, homologues, variants or complements and their encoded polypeptides. Also included are an expression vector comprising the polynucleotide operably linked to an expression control sequence; a host cell transformed or transfected with an expression vector of; an isolated antibody, or its antigen-binding fragment that specifically binds to the polypeptide; a method for detecting the presence of a cancer in a patient; a fusion protein comprising at least the polypeptide; an oligonucleotide that hybridises to the polynucleotide under moderately stringent conditions; a method for stimulating and/or expanding T cells specific for a tumour protein; an isolated T cell population comprising T cells prepared from the method of above; a composition comprising a first component consisting of carriers and immunostimulants, and a second component selected from the polynucleotides, proteins, antibodies, fusion proteins, T cell populations and antigen presenting cells expressing the polypeptide; methods for stimulating an immune response or treating cancer in a patient by administering the composition and diagnostic kits comprising at least one of the oligonucleotide of, or an antibody and a detection reagent consisting of a reporter group. The polypeptides and polynucleotides are useful as vaccines for the treatment or prevention of lung cancer, and for diagnosis and monitoring of such cancer. The polynucleotide, polypeptide and antigen presenting cells can be used to stimulate or expand T cells specific for a tumorous protein. The polynucleotides may be used as probes or primers for nucleic acid hybridisation, and in the preparation of ribozyme molecules for inhibiting expression of tumour polypeptides and proteins in tumour cells. The present sequence is one of the 183 lung cancer associated polynucleotides  
SQ Sequence 5739 BP; 1618 A; 1323 C; 1601 G; 1197 T; 0 U; 0 Other;  
Query Match 10.1%; Score 421.4; DB 6; Length 5739;  
Best Local Similarity 51.9%; Pred. No. 1.3e-101;

Matches 1236;		Conservative	0;	Mismatches 1076;	Indels	69;	Gaps	10;
QY	818	TACATCCATACCAAGCTTGAGGGACTTAAATTTTTCGGGTTCTCGTGGTCAAAAACAGACGC	877					
Db	2177	TGCACCCCTATCAAAATGGAGGGCCCTGAATTGTGTGCGTTCTCTCGGCTCAGGGCACTG	2236					
QY	878	ATGTAATCCTTGTGATGAAATGGGACTAGGCAAGACAAATTCAAAGCATTCGCCCTTTTAG	937					
Db	2237	ACACCATCTTTGGCTGATGATGGGCCCTTGGGAAAACGTGACAGACAGCAGTCTTCCTGT	2296					
QY	938	CTTCACCTTTTGGAGAG-----AACCTCATTCGGCATTTGGTAATTTGCTCCTCTATCGA	991					
Db	2297	ATTCCCTTTTACAAGAGGGTCAATTCCAAAGGCCCTTCTAGTAGAGCGCCCTCTTTCTA	2356					
QY	992	CTCTGCGTAACCTGGGAGAGAGAGTTTGGCCACATGGGCCCCCAACAGATGAACGTGGTTATGT	1051					
Db	2357	CCATCATCAACTGGGAGCGGGAGTTTGAAATGTGGGCTCCAGACATGTATGTCGTAACCT	2416					
QY	1052	ATTTTGGCACTGCGCAAGCTCGAGCAGTTTATCAGAGAAACATGAGTTTACTTATCGAAAG	1111					
Db	2417	ATGTGGGTGACAAAGACAGCCGTGCCATCATCGAGAGAATGAGTTCTCCTTTGAAGACA	2476					
QY	1112	ATCAAAAAAAGATCAAGAAAAAGAAATCTGGACAAATAAGTAGCGAAAGCAAGCAAAAA	1171					
Db	2477	ATGCCATTCGTGTGGCAAGAGGCCCTC-----CCGCATGAAGAAAGAGGCAT	2524					
QY	1172	GAATCAAGTTTGTATGTCTCTCACATCGTATGAGATGATCAACCTAGATTTCAGCAGTTC	1231					
Db	2525	CTGTGAAATTCATATGTGCTGTGACATCCATGAATTGATCACCATTCATGGCTATTT	2584					
QY	1232	TAAACCAATTAAGTGGGAGTGCAATGATTGTGATGAAGGTCAATCGACTGAAAAATAAGG	1291					
Db	2585	TGGGCTCTATTGATTGGGCCCTGCCCTCATCGTGGATGAAGCCCATCGGCTGAAGAACATC	2644					
QY	1292	ATTCAAAGCTGTTCTCTTCATTGACACAGTATTCAAAGTAACCACCGTATCTTCTGACAG	1351					
Db	2645	AGTCTAAGTTCTTCGGGTATTGAAATGGTTACTCACTCCAGCACAAAGCTGTGTGACTG	2704					
QY	1352	GAACACCACTTCAGAACAACTTGGATGAACTTTTCATGCTCATGCAATTTCTTGATCGCG	1411					
Db	2705	GGACACCATTACAAAAACAACTCGAAGAGTTGTTTCATCTGCTCAACTTCTCACCCCG	2764					
QY	1412	GGAAGTTTGGAACTTGGAGGAGTTCAGGAGGAGTTCAAAGATATTATCAAGAGGAGC	1471					
Db	2765	AGAGTTCCACAAATTTGGAAGGTTTTTTGGAGGAGTTGTTGACATTTGCCAAGGAGACC	2824					
QY	1472	AGATCTCAAGTTGCACAAAATGTTGGCTCCACATTTGCTCAGAAAGGTAAAAAAGACG	1531					
Db	2825	AGATAAAAAAACTGCATGACATGCTGGGGCCGACATGTTGGCGGGCTCAAAGCCGATG	2884					
QY	1532	TAAATGAAAGACATGCCCCCCEAAAAAGGAGCTCAATTTTTCGGTGTGATCTGAGCAGTCTGC	1591					
Db	2885	TGTTCAAGAACATGCCCCCTCCAAGACAGAACTAATGTGTGCGTGTGGAGCTGAGCCCCATGC	2944					
QY	1592	AGAAAGAAATATTACAAAGCTATTTTACCCGTAATTATCAAGTATTGA-----CAAAAAAGG	1648					
Db	2945	AGAAGAAATACTACAAGTACATCCTCACTCGAAATTTTGAAGCACTCAATGCCCCGAGGTG	3004					
QY	1649	GAGGTGCTCAAAATTTCCCTTAATAACATTTATGATGGAATTACGAAAAGTATGCTGCCATC	1708					
Db	3005	GTGGCAACCAGGTGTCTCTGCTGAAATGTGGTGATGGAATCTTAAGAAGTGTGCAACCATC	3064					
QY	1709	CTTA-----TATGCTAGAGGGTGTGAGCCAGTTATTACAGCGCAAAATGAAGCTT	1759					
Db	3065	CATACCTCTTCCCTGTGGCTGCAATGGAAGCTCTTAAGATGCCCTAATGGCATGTATGATG	3124					
QY	1760	TCAAAACAACCTTTTGGAGTCTTGTGGAAAGCTGCAACTTCTAGATAAAATGATGGTCAAAC	1819					
Db	3125	GCAGTGCCCTAATACAGAGCATCTGGGAAATTAATGCTGCTGCAGAAAATGCTCAAGAAC	3184					
QY	1820	TGAAAGAGCAAGGACACAGAGTCCCTAATATACACAGTTTCAGCATATGCTGGACTTAC	1879					
Db	3185	TTAAGGAGGGTGGGCATCGTGACTCATCTTTTCCAGATGACCAAGATGCTAGACCTGC	3244					

QY	1880	TTGAAGACTACTGTACCCATAAGAAATGGCAGTACGAGCGAATTCATGGAAGGTTGGCG	1939
Db	3245	TAGAGGATTTCTTGGAAACATGAAGGTTATAAATACGAACGCATCGATGGTGAATCACTG	3304
QY	1940	GAGCTGAGCGGCAAAATACGCATAGATCGGTTCAATGCCAAAAATCTAAACAAGTTTGT	1999
Db	3305	GGAACATGCGGCAAGAGAGGCCATTGACCGCTTCAATGCACCCGGTGCTCAGCAGTCTGTCT	3364
QY	2000	TTTTGCTCTCCACAAAGAGCTGGTGGCTTAGGAATAAATCTTGCAACGGCTGATACAGTAA	2059
Db	3365	TCCTGCTTTCACACTCGAGCTGGGGCCCTTGAATCAATCTGGCCACTGCTGACACAGTTA	3424
QY	2060	TCATTTATGACAGTGACTGGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC	2119
Db	3425	TTATCTATGACTCTGACTGGAACCCCCATAATGACATTCAGGCTTTAGCAGAGCTCACC	3484
QY	2120	GACTTGGCCAAAACAAATAAAGGTGATGATTTATAGGCTCATAAACCGAGGCACCATTTGAAG	2179
Db	3485	GGATTGGCAAAATAAAAAGGTATGATCTACCGTTTGTGACCCGTGCGTCAGTGGAGG	3544
QY	2180	AAAGGATGATGCAATTGACTAAAAAGAAAATGGTTCTAGAGCATCTTGTGTTGG-----	2234
Db	3545	AGCGCATCAGCAGGTGGCAAGAGAAAATGATGCTGACGCATCTAGTGTGCGGCCTG	3604
QY	2235	----GAAACTCAAAAACAAAAACATTAATCAGGAAGAGTTAGATGACATCATCAGGTATG	2290
Db	3605	GGCTGGGCTCCAAGACTGGATCTATGTCCAAAACAGGAGCTTGATGATATCCTCAAATTTG	3664
QY	2291	GATCAAAAGGAGCTTTTTCCTAGTGAAGATGATGAAGCAGGAAAAGTCTGGAATAATTCATT	2350
Db	3665	GCACCTGAGGAACTATTCAAGGATGAAGCCACTGATGGAGGAGAGACAACAAAGAGGGAG	3724
QY	2351	ATGATGATCGCGCTATAGACAAAATTCCTTGA-----TCGTGATCTCGTGAGGCAGAGG	2404
Db	3725	AAGATAGCAGTGTATCCACTACGATGATAAGGCCATTGAAACGCTGCTAGACCGTAACC	3784
QY	2405	AAGTCTCAGTGGATGATGAAGAGGAGAAATGGATTCTTTAAAGGCTTTCAGAGGTGGCTAAT	2464
Db	3785	AGGATGAGACTGAAGACACAGAAATTCAGGGGCATGAATGAATATTTGAGCTCATTCAAAG	3844
QY	2465	TTGAATATATAGATGAAAAATGAGGCAGCAGCATTTAGAGGCACAGAGAGTCGCTGAAA	2524
Db	3845	TGGCCAGTATGTGGTACGGGAAGAGAAAATGGGGAGGAAGAGAGGTAGAACGGGAAA	3904
QY	2525	GCAAACTTTCAGCAGGCAATCTGTATAGACCAAGTTATTGGGAAGAGTTGTTAAAAGATA	2584
Db	3905	TCATTAAACAGGAAGAAAGTGTGGATCCTGA---CTACTGGGAGAAATTCCTGCGGCACC	3961
QY	2585	AATTTGAGCTGCACCAAGGCTGAGGAGCTTAATGCTCTTTGGAAAAAGGAAGAGAGTCGCA	2644
Db	3962	ATTATGAGCAGCAGCAAGAAGATCTAGCCCGAAATCTGGGCAAGGAAAAAGAAATCCGTA	4021
QY	2645	AGCAGTTGGTATCCATTGAAG--AAGATGATCTTGTGTTTGGAAAGATGTGAGCTCTG	2701
Db	4022	AACAGGTCAACTACAATGATGGCTCCAGGAGGACCGAGATTGGCAGGACGACCAAGTCCG	4081
QY	2702	ATGGAGATGAAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGACAAGGAGTTC	2761
Db	4082	ACAACCAGTCCGATTACTCAGTGGCTTCAGAGGAAGGTGATGAAGACTTTGATGAACGTT	4141
QY	2762	AGACGGGTGACGGCCGTACAGAAGAAAGGTCGGGATTAATTTGGAACCAA-----	2812
Db	4142	CAGAAGCTCCCCTAGGCCCAGTCGTAAAGGCCCTGCGGAATGATAAAGATAAGCCATTGC	4201
QY	2813	CTCCGTTGATGGAAGGTGAGGGGAGATCTTTTCAGAGTACTGGGTTTCAACCAGAGTCAAA	2872
Db	4202	CTCCTCTGTGGCCCGTGTGGTGGGAATATTGAAGTACTTGGTTTAAATGCTCGTCAGC	4261
QY	2873	GGGCCATTTTGTACAGACTTTTGTATGAGGTATGG-----AGCTGGCAATTTTGATT	2923
Db	4262	GAAAGCCCTTCTTAATGCAATTATGCGATATGTTATGCCACCTCAGGATGCTTTTACTA	4321

Qy	2924	GGAAGGAGTTTGTTCCTCGCTTAAAGCAGAAGACCTTTGAAGAAAATAAATGAATATGGAA	2983
Db	4322	CCCAGTGGCTTTGAAGAGACCTGCGAGGCCAAATCAGAGAAAAGAGTTCAAGGCATATGTCT	4381
Qy	2984	TACTCTTCTTGAAGCACATTTGCTGAAGAAATAGACGAGAATTTCTCCAACCTTTTCAGATG	3043
Db	4382	CTCTTTTCATCGGCATTATGTGAGCCGGGGGAGATGGGGCTGAGACCTTTTGTGTGATG	4441
Qy	3044	GTGTGCCCCAAGGAAGGACTTAGAATAGAAGATGTTCTTAGTCAGAAATTGCTCTTCTGTATAC	3103
Db	4442	GTGTCCCCCGAAGAGGCCCTGTCTCGCCAGCATGTCTTACTAGAAATTGGTGTATGTCTTT	4501
Qy	3104	TAGTTCAGGAGAAAGGTGAAAATTTGTAGAAGATCATCCAGGG	3144
Db	4502	TGATTCGCAAGAGGTTTCAGGAGTTTGAACATGTTAATGGG	4542

RESULT 5	
ABL64410	
ID	ABL64410 standard; DNA; 6417 BP.
XX	
AC	ABL64410;
XX	
DT	15-MAY-2002 (first entry)
XX	
DE	Stomach cancer related gene sequence SEQ ID NO:2747.
XX	
KW	Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW	stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW	cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW	gene; ds.

PR	05-JUN-2000;	2000US-0209473P.
PR	05-JUN-2000;	2000US-0209531P.
PR	18-SEP-2000;	2000US-0233133P.
PR	18-SEP-2000;	2000US-02333617P.
PR	20-SEP-2000;	2000US-0234009P.
PR	20-SEP-2000;	2000US-0234034P.
PR	20-SEP-2000;	2000US-0234052P.
PR	22-SEP-2000;	2000US-0234509P.
PR	22-SEP-2000;	2000US-0234567P.
PR	25-SEP-2000;	2000US-0234923P.
PR	25-SEP-2000;	2000US-0234924P.
PR	25-SEP-2000;	2000US-0235077P.
PR	25-SEP-2000;	2000US-0235082P.
PR	25-SEP-2000;	2000US-0235134P.
PR	25-SEP-2000;	2000US-0235280P.
PR	26-SEP-2000;	2000US-0235637P.
PR	26-SEP-2000;	2000US-0235638P.
PR	27-SEP-2000;	2000US-0235711P.
PR	27-SEP-2000;	2000US-0235720P.
PR	27-SEP-2000;	2000US-0235840P.
PR	27-SEP-2000;	2000US-0235863P.
PR	28-SEP-2000;	2000US-0236028P.
PR	28-SEP-2000;	2000US-0236032P.
PR	28-SEP-2000;	2000US-0236033P.
PR	28-SEP-2000;	2000US-0236034P.
PR	28-SEP-2000;	2000US-0236109P.
PR	28-SEP-2000;	2000US-0236111P.
PR	29-SEP-2000;	2000US-0236842P.
PR	29-SEP-2000;	2000US-0236891P.
PR	02-OCT-2000;	2000US-0237172P.
PR	02-OCT-2000;	2000US-0237173P.
PR	02-OCT-2000;	2000US-0237278P.
PR	02-OCT-2000;	2000US-0237294P.

02-OCT-2000; 2000US-0237295P.  
02-OCT-2000; 2000US-0237316P.  
03-OCT-2000; 2000US-0237425P.  
03-OCT-2000; 2000US-0237598P.  
03-OCT-2000; 2000US-0237604P.  
03-OCT-2000; 2000US-0237606P.  
03-OCT-2000; 2000US-0237608P.  
01-NOV-2000; 2000US-0244867P.  
01-NOV-2000; 2000US-0245084P.  
XX  
PA (AVAL-) AVALON PHARM.  
XX  
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
PI Soppet DR, Weaver Z;  
XX  
DR WPI; 2002-188264/24.  
XX  
PT Screening for anti-neoplastic agent involves exposing cells to a chemical  
PT agent to be tested for anti-neoplastic activity, and determining a change  
PT in expression of a gene of a signature gene set.  
XX  
PS Claim 1; SEQ ID NO 2747; 44pp; English.  
XX  
CC The present invention describes a method (M1) for screening for an anti-  
CC neoplastic agent. The method involves exposing cells to a chemical agent  
CC to be tested for anti-neoplastic activity, determining a change in  
CC expression of at least one gene (I) of a signature gene set, where (I)  
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664  
CC to ABL70110), or is at least 95% identical to (S), where a change in  
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
CC activity and can be used in gene therapy. M1 can be used for screening an  
CC anti-neoplastic agent, and can be used for producing a product which is  
CC the data collected with respect to the anti-neoplastic agent as a result  
CC of M1, and the data is sufficient to convey the chemical structure and/or  
CC properties of the agent. M1 can be used in the treatment of cancer such  
CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,  
CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell  
CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous  
CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's  
CC tumour  
XX  
SQ Sequence 6417 BP: 1774 A; 1494 C; 1784 G; 1365 T; 0 U; 0 Other;

QY 1172 GAATCAAGTTTGATGTCTCTCCTCACATCGTATGAGATGATCAACCTAGATTTCAGCAGTTT



Db 2614 CTGTGAAATTCCATGTGCTGTGACATCCTATGAATTGATCACCAATTGACATGGCTATTT 2673

QY 1232 TAAAAACCAATTAAAGTGGAGTGCATGATTGTTGATGAAGGTCAATCGACTGAAAAATAAGG 1291

Db 2674 TGGGCTCTATTGATTGGGCTGCCTCATCGTGGATGAAGCCCATCGGCTGAAGAACAATC 2733

QY 1292 ATTCAAAGCTGTTCTCTTCATTGACACAGTATTCAAAGTAACCAACCGTAATCTTCTGACAG 1351

Db 2734 AGTCTAAGTTCCTCCGGGTATTGAATGGTTACTCACTCCAGCACAAAGCTGTGCTGACTG 2793

QY 1352 GAAACACCACCTTCAGAAACAACCTTGGATGAACCTTTTCATGCTCATGCAATTTCTTGTATGCGG 1411

Db 2794 GGACACCATTACAAAACAATCTGGAAGAGTTGTTTCATCTGCTCAAACITTTCTCACCCCGG 2853

QY 1412 GGAAGTTTGGAGTTTGAGGAGTTCAGGAGGAGTTCCAGAGGAGTTCAAAGATATTAAATCAAGAGGAGC 1471

Db 2854 AGAGGTTCCACAATTTGGAAGGTTTTTTGGAGGAGTTTGCTGACATTGCCAAGGAGGACC 2913

QY 1472 AGATCTCAAGGTTGCACAAAAATGTTGGCTCCACATTTGCTCAGAAAGGTAAAAAAAAGACG 1531

Db 2914 AGATAAAAAAACTGCATGACATGCTGGGCCGCACATGTTGCGCGGCTCAAAAGCCGATG 2973

QY 1532 TAATGAAAGACATGCCCCCAAAAAAGGAGCTCAATTTGCGTGTGATCTGAGCAGTCTGC 1591

Db 2974 TGTTCAAGAACATGCCCTCCAAGACAGAACTAATTTGTGCGTGTGAGCTGAGCCCTATGC 3033

QY 1592 AGAAAGAAATATTACAAAGCTATTTTACCCCGTAATTATCAAGTATTGA---CAAAAAAGG 1648

Db 3034 AGAAGAAATACTACAAGTACATCCTCACTCGAAATTTTGAAGCACTCAATGCCCGAGGTG 3093

QY 1649 GAGGTGCTCAAAATTTCCCTTAATAACATTTATGATGGAATTACGAAAAGTATGCTGCCATC 1708

Db 3094 GTGGCAACCAAGGTGCTCTGCTGAAATGTGGTGATGGATCTTAAGAAGTGCTGCAACCATC 3153

QY 1709 CTTA-----TATGCTAGAGGGTGTGAGCCAGTTATTCAAGCGCAATGAAGCTT 1759

Db 3154 CATACCTCTTCCCTGTGGTGCAAATGGAAGCTCCTAAGATGCCTAATGGCATGTATGATG 3213

QY 1760 TCAAAACAACCTTTTGGAGTCTTGTGAAAAGCTGCAACTTCTAGATAAAATGATGGTCAAAC 1819

Db 3214 GCAGTGCCCTAATCAGAGCATCTGGGAAATTAATGCTGCTGCAGAAAAATGCTCAAGAACC 3273

QY 1820 TGAAGAGCAAGGACACAGAGTCCTAATATACACACAGTTTCAGCATATGCTGACCTTAC 1879

Db 3274 TTAAGGAGGTGGGCATCGTGTACTCATCTTTTCCAGATGACCAAGATGCTAGACCTGC 3333

QY 1880 TTGAAGACTACTGTACCCATAAGAAATGGCAGTACGAGCGAAATTGATGGAAGGTTGGCG 1939

Db 3334 TAGAGGATTTCTTGGAAACATGAAGGTTATAAATACGAACGCATCGATGGTGGAAATCACTG 3393

QY 1940 GAGCTGAGCGGCAAAATCAGCATAGATCGGTTCAATGCCAAAAAATTTCAACAGTTTTGTT 1999

Db 3394 GGAACATGCGGCAAGAGGCCATTGACCGCTTCAATGCAACCGGGTCTCAGCAGTTCTGCT 3453

QY 2000 TTTTGTCTCCACAAGAGCTGGTGGCTTAGGAATAATCTTGCAACGGCTGATACAGTAA 2059

Db 3454 TCTTGCTTTCACCTCGAGCTGGGGGCCCTTGGAAATCAATCTGGCCACTGCTGACACAGTTA 3513

QY 2060 TCATTTATGACAGTGACTGGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC 2119

Db 3514 TTATCTATGACTCTGACTGGAACCCCAATAATGACATTCAGGCCTTTAGCAGAGCTCACC 3573

QY 2120 GACTTGGCCAAAACAATAAGGTGATGATTTATAGGCTCATAAACCGAGGCCACCATTTGAAG 2179

Db 3574 GGATTGGCCAAAATAAAAAGGTAATGATCTACCCGGTTGTGACCCCGTGCCTCAGTGGAGG 3633

QY 2180 AAAGGATGATGCAATTGACTAAAAAGAAAAATGGTTCTAGAGCATCTTGTGTTGTG----- 2234

Db 3634 AGCGCATCAGCAGGTGGCAAGAAGAAAAATGATGCTGACGCATCTAGTGGTGGCGCCTG 3693

QY 2235 ----GAAACTCAAAACACAAAACATTAATCAGGAAGAGTTAGATGACATCATCAGGTATG 2290

Db 3694 GGCTGGGCTCCAAGACTGGATCTATGTGTCCAAACAGGAGGCTTGATGATATCCTCAAATTTG 3753

QY 2291 GATCAAAGGAGCTTTTTTGCTAGTGAAGATGATGAAGCAGGAAAGTCTGAAAAAATTCATT 2350

Db 3754 GCACTGAGGAACCTATTTCAAGGATGAAGCCACTGATGGAGGAGGACAAACAAGAGGGAG 3813

QY 2351 ATGATGATCGGGCTATAGACAAATTCCTTGA-----TCGTGATCTCGTGGAGGCAGAGG 2404

Db 3814 AAGATAGCAGTGTATTCCACTACGATGATAAGGCCATTGAACGGCTGTAGACCGTAACC 3873

QY 2405 AAGTCTCAGTGGATGATGAAGAGGAGAAATGGATTCTTAAAGGCTTTCAGGTTGGCTAATT 2464

Db 3874 AGGATGAGACTGAAGACACAGAAATTCAGGGGCATGAATGAATATTTGAGCTCATTCAAAG 3933

QY 2465 TTGAATATATAGATGAAAAATGAGGCAGCAGCATTAGAGGCACAGAGATCGCTGTGAAA 2524

Db 3934 TGGCCCACTGTGGTACGGGAAGAAAAATGGGGAGGAAGAGGAGGTAGAACGGGAAA 3993

QY 2525 GCAAACTTTCAGCAGGCAATTCGTATAGAGCAAGTTATTGGGAAGAGTTGTTAAAAAGATA 2584

Db 3994 TCATTAAACAGGAAGAAAGTGTGGATCCTGA---CTACTGGGAGAAATTCGTCGGGCACC 4050

QY 2585 AATTGAGCTGCACCAAGGCTGAGGAGCTTAATGCTCTTGGAAAAAGGAAGAGATCGCA 2644

Db 4051 ATTTATGAGCAGCAGCAAGAAGATCTAGCCCCGAAATCTGGGCAAGGAAGAAAGATCCGTA 4110

QY 2645 AGCAGTTGGTATCCATTGAAG---AAGATGATCTTGTGTTTGGAAAGATGTGAGCTCTG 2701

Db 4111 AACAGGTCAACTACAATGATGGCTCCAGGAGGACCGAGATTGGCAGGACGACCAGTCCG 4170

QY 2702 ATGGAGATGAAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGACAAGGAGTTC 2761

Db 4171 ACNACCAGTCGGATTACTCAGTGGCTTCAGAGGAAGGTGATGAAGACTTTGATGAACGTT 4230

QY 2762 AGACGGGTCGACGGCCGTACAGAAAGAAAGGTCGCGATAAATTTGGAACCAA----- 2812

Db 4231 CAGAAGCTCCCCGTAGGCCCAGTCGTAAGGGCCTGCGGAATGATAAAGATAAGCCATTGC 4290

QY 2813 CTCCGTTGATGGAAGGTGAGGGGAGATCTTTCAGAGTACTGGGTTTCAACCCAGAGTCAAA 2872

Db 4291 CTCCTCTGTGGCCCCGTGTTGGTGGGAATATTGAAGTACTTGGTTTTAATGCTCGTCAGC 4350

QY 2873 GGGCCATTTTGTACAGACTTTTGATGAGGTATGG-----AGCTGGCAATTTTGATT 2923

Db 4351 GAAAAGCCTTCTTAATGCAATTTATGCGATATGGTATGCCACCTCAGGATGCTTTTACTA 4410

QY 2924 GGAAGGAGTTTGTTCCTCGCTTAAAGCAGAAACCTTTGAAGAATAAATGAATATGGAA 2983

Db 4411 CCAGTGGCTTGTAAAGAGACCTCGCAGGCAAAATCAGAGAAAGAGTTCAAGGCATATGTCT 4470

QY 2984 TACTCTTCTGAAGCACATTTGCTGAAGAAATAGACGAGAAATTCCTCAACCTTTTTCAGATG 3043

Db 4471 CTCTTTTCATGCGGCATTTATGTGAGCCGGGCAGATGGGCTGAGACCTTTGCTGATG 4530

QY 3044 GTGTGCCCCAAGGAAGGACTTAGAATAGAAAGATGTTCTAGTCAGAATTCCTCTTGATAC 3103

Db 4531 GTGTCCCCCGAAGAGCCCTGTCTCGCCAGCATGTCTTACTAGAAATGSGTGTATGTCTT 4590

QY 3104 TAGTTCAGGAGAGGTGAAATTTGTAGAAAGATCATCCAGGG 3144

Db 4591 TGATTCCGAAGAAGGTTTCAGGAGTTTGAACATGTTAATGGG 4631

RESULT 6  
ABX04169  
ID ABX04169 standard; cDNA; 6417 BP.  
XX

AC ABX04169;

XX  
DT 10-JAN-2003 (first entry)

XX  
DE Human mRNA differentially expressed in mesenchymal cells #16.

XX  
KW Human; ss; gene; skeletal growth; cartilage degeneration disorder;



Db	3454	TCTTGCTTTCCACTCTGAGCTGGGGCCCTTGGAAATCAATCTGGCCACTGCTGACACAGTTA	3513
QY	2060	TCATTTATGACAGTGAATCGGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC	2119
Db	3514	TTATCTATGACTCTGACTGGAACCCCAATAATGACATTCAGGCCTTTAGCAGAGCTCACC	3573
QY	2120	GACTTGGCCAAACAATAAGGTGATGATTTATAGGCTCATAAACCGAGGCACCATGGAAG	2179
Db	3574	GGATTGGGCAAAATAAAAAGGTAATGATCTACCGGTTTGTGACCCGTCGCTCAGTGGAGG	3633
QY	2180	AAAGGATGATGCAATTGACTAAAAAGAAAATGGTTCTAGAGCATCTTGTGTTTGG-----	2234
Db	3634	AGCGCATCACGCAAGTGGCAAGAAGAAAATGATCTGACGCATCTAGTGTGTCGGCCTG	3693
QY	2235	----GAAACTCAAAACACAAAAACATTAATCAGGAAGAGTTAGATGACATCATCAGGTATG	2290
Db	3694	GGCTGGCTCCAAGACTGGATCTATGTCCAAACAGAGAGCTTGATGATATCCTCAAAATTG	3753
QY	2291	GATCAAAAGGAGCTTTTGTGCTAGTGAAGATGATGAAGCAGGAAAAGTCTTGAAAAATTCATT	2350
Db	3754	GCACTGAGGAACATAATCAAGGATGAAGCCACTGATGAGGAGGAGACAACAAGAGGGAG	3813
QY	2351	ATGATGATCGGGCTATAGACAAATTGCTTGA-----TCGTGATCTCGTGGAGGCAGAGG	2404
Db	3814	AAGATAGCAGTGTATCCACTACGATGATAAGGCCATTGAACGGCTGCTAGACCGTAACC	3873
QY	2405	AAGTCTCAGTGGATGAAGAGGAGAAATGGATTCTTAAAGGCTTTCAAGGTGGCTAATT	2464
Db	3874	AGGATGAGACTGAAGACACAGAAATTGCAGGGCATGAATGAATATTTGAGCTCATTCAAAG	3933
QY	2465	TTGAATATATAGATGAAAATGAGGCAGCAGCATTTAGAGGCACAGAGAGTCGCTGCTGAAA	2524
Db	3934	TGGCCAGTATGTGTTACGGGAAGAAGAAATGSSGAGGAAGAGGAGGTAGAACGGGAAA	3993
QY	2525	GCAAATCTTCAGCAGGCAATCTGTATAGAGCAAGTTATTGGGAAGAGTTGTTAAAAAGATA	2584
Db	3994	TCATTAAACAGGAAGAAAGTGTGGATCCTGA---CTACTGGGAGAAAATTGCTGCGGCACC	4050
QY	2585	AATTTGAGCTGCACCGCTGAGGAGCTTAATGCTCTTGGAAAAAGGAAGAGAGTCCGA	2644
Db	4051	ATTATGAGCAGCAGCAAGAAGATCTAGCCCAGAAATCTGGGCAAGGAAAAAGAAATCCGTA	4110
QY	2645	AGCAGTTGGTATCCATTGAAG--AAGATGATCTTGTGGTTTGGAAAGATGTAGCTCTG	2701
Db	4111	AACAGGTCAACTACAATGATGGCTCCAGGAGGACCAGAGATTGGCAGGACGACCATCCG	4170
QY	2702	ATGGAGATGAAAGTTATGAAGCTGAGTCAACAGATGTTGAAGCAGCAGGACAAGGAGTTC	2761
Db	4171	ACAACAGTCCGATTACTCAGTGGCTTCAGAGGAAGGTGATGAAGACTTTGATGAACGTT	4230
QY	2762	AGACGGTTCGACGGCCGTAAGAAAGAAAGGTTCGCGATAATTTGGAACCA-----	2812
Db	4231	CAGAAAGTCCCCGTAGGCCAGTCGTAAGGGCCCTGCGGAATGATAAAGATAAGCCATTGC	4290
QY	2813	CTCCGTTGATGGAAGGTGAGGGGAGATCTTTCAGAGTACTGGGTTTCAACCAGAGTCAAA	2872
Db	4291	CTCCTCTGTTGGCCCGTGTGGTGGGAATATTGAAGTACTTGGTTTAAATGCTCGTCAGC	4350
QY	2873	GGGCCATTTTGTACAGACTTTTGATGAGGTATGG-----AGCTGGCAATTTTGATT	2923
Db	4351	GAAAAGCCTTTCTTAATGCAATTATGCGATATGGTATGCCACCTCAGGATGCTTTTACTA	4410
QY	2924	GGAAGGAGTTTGTTCCTTCCTTAAAGCAGAAGACCTTTGAAGAAAATAAATGAATATGGAA	2983
Db	4411	CCCAGTGGCTTGTAAAGAGACTGCGAGGCCAAATCAGAGAAAAGAGTTCAAGGCATATGTCT	4470
QY	2984	TACTCTTTGAAGCACATTGCTGAAGAAAATAGACGAGAAATTCTCCAACCTTTTCAGATG	3043
Db	4471	CTCTTTTCATGCGGCATTTATGTAGCCGGGGCAGATGGGGCTGAGACCTTTTGCTGATG	4530
QY	3044	GTGTGCCCAAGGAAGGACTTAGAATAGAAGATGTTCTAGTCAGAAATTGCTCTTCTGATAC	3103

Db	4531	GTGTCCCGCAGAGGCCTGTCTCGCCAGCATGTCTTACTAGAAATTGGTGTATGTCTT	4599
Qy	3104	TAGTTCAGGAGAAGGTGAAATTTGTAGAAGATCATCCAGGG	3144
Db	4591	TGATTCGCAAGAAGGTTTCAGGAGTTTGAACATGTTAATGGG	4631
RESULT 7			
ABK84696			
ID	ABK84696 standard; cDNA; 6417 BP.		
XX	XX		
AC	AC		
XX	XX		
DT	14-AUG-2002 (first entry)		
DE	Human cDNA differentially expressed in granulocytic cells #1267.		
XX	XX		
KW	Human; ss; granulocytic cell; DNA chip; bacterial infection;		
KW	viral infection; parasitic infection; protozoal infection;		
KW	fungal infection; sterile inflammatory disease; psoriasis;		
KW	rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;		
KW	cardiac reperfusion injury; renal reperfusion injury; ARDS;		
KW	adult respiratory distress syndrome; inflammatory bowel disease;		
KW	Crohn's disease; ulcerative colitis; periodontal disease;		
KW	granulocyte activation; chronic inflammation; allergy.		
XX	XX		
OS	Homo sapiens.		
XX	XX		
PN	WO200228999-A2.		
XX	XX		
PD	11-APR-2002.		
XX	XX		
PF	03-OCT-2001; 2001WO-US030821.		
XX	XX		
PR	03-OCT-2000; 2000US-0237189P.		
XX	XX		
PA	(GENE-) GENE LOGIC INC.		
XX	XX		
PI	Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;		
XX	WPI; 2002-435328/46.		
DR	XX		
XX	XX		
PT	Detecting granulocyte activation by detecting differential expression of		
PT	genes associated with granulocyte activation, which serves as diagnostic		
PT	markers that is useful for monitoring disease states and drug toxicity.		
XX	XX		
PS	Claim 1; SEQ ID NO 1267; 114pp; English.		
XX	XX		
CC	The invention relates to detecting (M1) granulocyte (GC) activation		
CC	(GCA), by detecting the level of expression of gene(s) (Gs) identified by		
CC	DNA chip analysis as given in the specification, and comparing the		
CC	expression level to an expression level in an unactivated GC, where		
CC	differential expression of Gs is indicative of GCA. Also included are		
CC	modulating (M2) GA by contacting GC with an agent that alters the		
CC	expression of at least one gene in Gs; (2) screening (M3) for an agent		
CC	capable of modulating GCA or an inflammation (especially chronic) in a		
CC	tissue, an allergic response in a subject, exposure of a subject to a		
CC	pathogen or sterile inflammatory disease using the gene expression		
CC	profile; (3) detecting (M4) an inflammation (especially chronic) in a		
CC	tissue, an allergic response in a subject, exposure of a subject to a		
CC	pathogen or sterile inflammatory disease, by detecting the level of		
CC	expression in a sample of the tissue of gene(s) from Gs, where the level		
CC	of expression of the gene is indicative of inflammation; (4) treating		
CC	(M5) an inflammation (especially chronic) or in a tissue, an allergic		
CC	response in a subject, exposure of a subject to a pathogen or sterile		
CC	inflammatory disease, by contacting a tissue having inflammation with an		
CC	agent that modulates the expression of gene(s) from Gs in the tissue. M1		
CC	is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful		
CC	for screening an agent capable of modulating GCA preferably in an		
CC	inflammation in a tissue; M4 is useful for detecting an inflammation		
CC	(especially chronic) in a tissue, an allergic response in a subject,		
CC	exposure of a subject to a pathogen or sterile inflammatory disease (e.g.		
CC	psoriasis, rheumatoid arthritis, glomerulonephritis, asthma thrombosis,		





Db 4171 ACAACAGTCCGATTACTCAGTGGCTTCAGAGGAAGGTGATGAAGACTTTTGATGAACGTT 4230  
QY 2762 AGACGGTCGACGGCGGTACAGAAAGAAAGGTCGCGATAATTTGGAAACAA----- 2812  
Db 4231 CAGAAAGCTCCCCGTAGGCCCAGTCGTAAAGGGCTCGGGAATGATAAAGATAAGCCATTGC 4290  
QY 2813 CTCGGTTGATGGAAGGTGAGGGGAGATCTTTCAGAGTACTGGGTTTCAACCAGAGTCAAA 2872  
Db 4291 CTCCTCTGTTGGCCCGTGTGTTGGTGGGAATATTGAAGTACTTGGTTTTAATGCTCGTCAGC 4350  
QY 2873 GGGCCATTTTGTACAGACTTTTGATGAGGTATGG-----AGCTGGCAATTTTGATT 2923  
Db 4351 GAAAAGCCTTTCTTAATGCAATTATGCGATATGTTATGCGATATGTTATGCCACCTCAGGATGCTTTACTA 4410  
QY 2924 GGAAGGAGTTTGTTCCTCGCTTAAAGCAGAGACCTTTGAAGAAATAAATGAATATGGAA 2983  
Db 4411 CCCAGTGGCTTGTAAAGAGACCTGCGAGGCAATCAGAGAAAGAGTTCAGGCATATGTCT 4470  
QY 2984 TACTCTCTTGAAGCACATTGCTGAAGAAATAGACGAGAATTCCTCCAACCTTTTCAGATG 3043  
Db 4471 CTCTTTTCATCGGGCAATTATGTGAGCCGGGGCAGATGGGGCTGAGACCTTTGCTGATG 4530  
QY 3044 GTGTGCCCAAGGAAGACTTAGAATAGAAGATGTTCTAGTCAGAAATTGCTCTTCTGATAC 3103  
Db 4531 GTGTCCCCGAGAAGGCCTGTCTCGCCAGCATGTCCTTACTAGAAATTGGTGTATGTCTT 4590  
QY 3104 TAGTTCAGGAGAAAGTGAAATTTGTAGAAGATCATCCAGGG 3144  
Db 4591 TGATTCGCAAGAAGGTTCAGGAGTTTGAACATGTTAATGGG 4631

RESULT 8

AAD54631  
ID AAD54631 standard; DNA; 6417 BP.  
XX AAD54631;  
AC AAD54631;  
XX 26-JUN-2003 (first entry)  
DT Human chromodomain helicase DNA binding protein (CHD) encoding DNA #6.  
DE Human; p53 pathway; therapeutic; angiogenic disorder; apoptotic disorder;  
XX Human; p53 pathway; therapeutic; angiogenic disorder; apoptotic disorder;  
KW chromodomain helicase DNA binding protein; CHD; cancer; gene therapy;  
KW cell proliferative disorder; chromatin organisation modifier domain;  
KW cytosstatic; ds.  
XX Homo sapiens.  
OS WO200298899-A2.  
XX 12-DEC-2002.  
XX 03-JUN-2002; 2002WO-US017466.  
XX 05-JUN-2001; 2001US-0296076P.  
PR 10-OCT-2001; 2001US-0328605P.  
PR 22-OCT-2001; 2001US-0338733P.  
PR 15-FEB-2002; 2002US-0357253P.  
PR 15-FEB-2002; 2002US-0357600P.  
XX (EXEL-) EXELIXIS INC.  
XX Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;  
PI Lioubin MN;  
XX WPI; 2003-156840/15.  
XX Identifying a candidate p53 pathway-modulating agent as therapeutic  
PT targets for disorders related to defective p53 function e.g. cancer by  
PT contacting an assay system having purified CHD polypeptide or nucleic  
PT acid, with a test agent.  
XX

PS Disclosure; Page 59-62; 278pp; English.  
XX The present invention relates to a method for identifying candidate p53  
CC pathway modulating agents. The method involves contacting an assay system  
CC comprising purified chromatin organisation modifier (chromo) domain  
CC helicase DNA binding proteins (CHD), nucleic acids, their functionally  
CC active fragments or derivatives, with a test agent under conditions  
CC where, but for the presence of the test agent, the system provides a  
CC reference activity. The methods are useful for identifying modulators of  
CC the p53 pathway as therapeutic targets for disorders associated with  
CC defective p53 function, such as angiogenic disorders, apoptotic disorders  
CC or cell proliferative disorders, e.g. cancer. The modulators are useful  
CC as research reagents, diagnostics and therapeutics. The invention is also  
CC useful in gene therapy. The present sequence is human CHD DNA  
XX  
SQ Sequence 6417 BP; 1774 A; 1494 C; 1784 G; 1365 T; 0 U; 0 Other;  
Query Match 10.1%; Score 421.4; DB 7; Length 6417;  
Best Local Similarity 51.9%; Pred. No. 1.4e-101;  
Matches 1236; Conservative 0; Mismatches 1076; Indels 69; Gaps 10;  
QY 818 TACATCCATACCAGCTTGAGGGACTTAATTTTTTGGGTTCTCGTGGTCAAAACAGACGC 877  
Db 2266 TGCACCCCTATCAAAATGGAGGCCCTGAATTGGTTGCGCTTCTCTGGGCTCAGGGCACTG 2325  
QY 878 ATGTAATCCTTGCTGATGAAATGGGACTAGGCAAGACAATTCAAAGCATTGCCCTTTTAG 937  
Db 2326 ACACCATCTTGGCTGATGAGATGGGCTTGGGAAACTGTACAGACAGCAGTCTTCCTGT 2385  
QY 938 CTTCACTTTTGTAGGAG-----AACCTCATTCGCGATTGTTGGFAATTGCTCCTCTATCGA 991  
Db 2386 ATTCCCTTTACAAGGAGGGTCATTCCAAAGGCCCTTCTCTAGTGGCGCCCTCTTTCTA 2445  
QY 992 CTCTGCGTAACTGGGAGAGAGTGTGCCACATGGGCCCCACAGATGAACGTGGTTATGT 1051  
Db 2446 CCATCATCAACTGGGAGCGGGAGTTGAAATGTGGCTCCAGACATGTATGCGTAACCT 2505  
QY 1052 ATTTTGGCACTGCCGCAAGCTCGAGCAGTTATCAGAGAACATGAGTTTACTTATCGAAAG 1111  
Db 2506 ATGTGGGTGACAAGGACAGCCGTGCCATCATCCGAGAGATGAGTTCTCCTTTGAAGACA 2565  
QY 1112 ATCAAAAAAAGATCAAGAAAAAGAAAATCTGGACAAATAAGTAGCGAAAGCAAAAAA 1171  
Db 2566 ATGCCATTCTGGTGGCAAGAGAGGCTC-----CCGATGAAGAAAGAGGCAT 2613  
QY 1172 GAATCAAGTTTGATGTCTCTCACATCGTATGAGATGATCAACCTAGATTCAGCAGTTC 1231  
Db 2614 CTGTGAAATTCATGTGCTGATGATCCTATGAATTGATCACCATTGACATGGCTATTT 2673  
QY 1232 TAAACCAATTAAGTGGGAGTGCATGATTGTTGATGAAGGTGCATCGACTGAAAAATAAGG 1291  
Db 2674 TGGGCTCTATTGATTGGGCTGCTCATCGTGGATGAAGCCCATCGGCTGAAGAACAATC 2733  
QY 1292 ATTCAAAGCTGTTCTCTTCATTGACACAGTATTCAAGTAACCAACCGTATTCTTCTGACAG 1351  
Db 2734 AGTCTAAGTTCTTCGGGTATTGTAATGGTTACTCACTCCAGCACAAAGCTGTTGCTGACTG 2793  
QY 1352 GAACACCACTTCAGAAACAACCTTGGATGAACCTTTTCATGCTCATGCAATTTCTTGATGCGG 1411  
Db 2794 GGACACCAATTACAAAACAATCTGGAAGAGTTGTTTCATCTGCTCAACTTCTCACCCCG 2853  
QY 1412 GGAAGTTTGGAAAGTTTGGAGGAGTTCAGGAGGAGTTCAAAGATATTAAATCAAGAGGAGC 1471  
Db 2854 AGAGGTTCCACAATTTGGAAGGTTTGTGGAGGAGTTTGTGACATTGCCAAGGAGGACC 2913  
QY 1472 AGATCTCAAGGTTGCACAAAAATGTTGGCTCCACATTTGCTCAGAAGGGGTAAAAAAGACG 1531  
Db 2914 AGATAAAAAAAACTGCATGACATGTCTGGGCGCGCACATGTTGCGGGGCTCAAAGCCGATG 2973  
QY 1532 TAATGAAAGACATGCCCCCCCCAAAAAGGAGCTCATTTTTCGCTGTTGATCTGAGCAGTCTGC 1591  
Db 2974 TGTTCAAGAACATGCCCTCCAAGACAGAACTAATTGTGCGTGTGAGCTGAGCCCTATGC 3033

QY	1592	AGAAAGAATATTACAAAGCTATTTTACCCGTAATTAATCAAGTATTGA---	CAAAAAAGG	1649
DB	3034	AGAAAGAAATACTACAAGTACATCCTCACTCGAAATTTGAAGCACTCAATGCGCGAGGTG		3093
QY	1649	GAGGTGCTCAAAATTTCCCTTAATAACATTAATGATGGAATACGAAAAAGTATGCTGCCATC	1708	
DB	3094	GTGGCAACCAGGTGTCTCTGCTGAATGTGGTGATGGATCTTAAGAAGTGTGCAACCATC	3153	
QY	1709	CTTA-----TATGCTAGAGGGTGTGAGCCAGTTATTCAACGCGCAATGAAGCTT	1759	
DB	3154	CATACCTCTTCCCTGTGGCTGCAATGGAAGCTCCTAAGATGCCTAATGGCATGTATGATG	3213	
QY	1760	TCAAAACAACTTTTGGAGTCTTTGTGGAAAGCTGCAACTTCTAGATAAAATGATGGTCAAAC	1819	
DB	3214	GCAGTGGCCCTAATCAGAGCATCTGGGAAATATTGCTGTGCAGAAAAATGCTCAAGAAC	3273	
QY	1820	TGAAAGAGCAAGGACACAGAGTCTTAATAATACACACAGTTTTCAGCATATGCTGGACTTAC	1879	
DB	3274	TTAAGGAGGGTGGGCATCGTGACTCATCTTTTCCAGATGACCAAGATGTAGACCTGC	3333	
QY	1880	TTGAAGACTACTGTACCCATAAGAAATGGCAGTACGAGCGAAATTGATGGAAGGTTGGCG	1939	
DB	3334	TAGAGGATTTCTTGGAAACATGAAGGTTATAATACGAACGCATCGATGGTGAATCACTG	3393	
QY	1940	GAGCTGAGCGGCAATACGCATAGATCGGTTCAATGCCCCAAAAATTTCTAACAGTTTGT	1999	
DB	3394	GGAACATGCGGCAAGAGGCCATTGACCGCTTCAATGACCGGTTGCTCAGCAGTTCTGCT	3453	
QY	2000	TTTTTGCTCTCCACAAGAGCTGGTGGCTTAGGAATAAATCTTGCAACGGCTGATACAGTAA	2059	
DB	3454	TCTTGCTTTCCACTCGAGCTGGGGCCCTTGGAAATCAATCTGGCCACTGCTGACACAGTTA	3513	
QY	2060	TCATTTATGACAGTGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC	2119	
DB	3514	TTATCTATGACTCTGACTGGAACCCCAATAATGACATTCAGGCCCTTTAGCAGAGCTCACC	3573	
QY	2120	GACTTGGCCAAACAATAAGGTGATGATTTATAGGCTCATAAACCGAGGCCAATTTGAAG	2179	
DB	3574	GGATTGGGCAAAATAAAAGGTAATGATCTACCGGTTGTGACCCGTTGCTCAGTGGAGG	3633	
QY	2180	AAAGGATGATGCAATTGACTAAAAAGAAATGGTTCTAGAGCATCTTGTGTGTGG---	2234	
DB	3634	AGCGCATCAGCAGGTGGCAAGAAAGAAATGATGCTGACGCACTCTAGTGGTGGCGCTG	3693	
QY	2235	---GAAACTCAAAACACAAAAACATTAAATCAGGAAGAGTTAGATGACATCATCAGGTATG	2290	
DB	3694	GGCTGGGCTCCAAGACTGGATCTATGTCCAACACGAGCTTGATGATATCTTCAAAATTTG	3753	
QY	2291	GATCAAAAGGAGCTTTTGTGCTAGTGAAGATGATGAACGAGGAAAGTCTTGAAAAATTCATT	2350	
DB	3754	GCACTGAGGAACATTCAAGGATGAAGCCACTGTAGAGGAGGAGACAAACAAAGAGGGAG	3813	
QY	2351	ATGATGATGCGGCTATAGACAAATTTGCTTGA-----TCGTGATCTCGTGGAGGCAGAGG	2404	
DB	3814	AAGATAGCAGTGTATTCCACTACGATGATAAGGCCATTGAACGGCTGTAGACCGTAAAC	3873	
QY	2405	AAGTCTCAGTGGATGAAGAGGAGAAATGGATTCTTAAAGGCTTTCAAGGTGGCTAATT	2464	
DB	3874	AGGATGAGACTGAAGACACAGAATTCGAGGGCATGAATGAATATTTGAGCTCATTCAAAG	3933	
QY	2465	TTGAATATATAGATGAATAATGAGGCAGCAGCATTAGAGGCAACAGAGATCGCTGTGANA	2524	
DB	3934	TGGCCCAGTATGTGGTACGGGAAGAAGAAATGGGGAGGAAGAGGAGGTAGAACGGGAAA	3993	
QY	2525	GCAAAATCTTCAGCAGGCAATCTGTAGAGCAAGTTATTGGGAAGAGTTGTTAAAAAGATA	2584	
DB	3994	TCATTAAACAGGAAGAAAGTGTGGATCCTGA--CTACTGGGAGAAATTTGCTGGGCACC	4050	
QY	2585	AATTTGAGCTGCCACCGCTGAGGAGCTTAATGCTCTTGGAAAAAGGAAGAGAACTCGCA	2644	
DB	4051	ATTATGAGCAGCAGCAAGAAGATCTAGCCCCGAAATCTGGGCAAAAGGAAAAAGAAATCCGTA	4110	
QY	2645	AGCAGTTGGTATCCATTGAAG---AAGATGATCTTGTGGTTTGGAAAGATGTAGCTCTG	2701	

Db	4111	AACAGGTCAACTACAATGATGGCTCCCGAGGACCGAGATTGGCAGAGACGACCGAGTCCG	4170
QY	2702	ATGGAGATGAAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGACAAGGAGTTT	2761
Db	4171	ACAACCAAGTCCGATTACTCAGTGGCTTCAGAGGAAGGTGATGAAGACCTTTGATGAACGTT	4230
QY	2762	AGACGGTCCGACGGCCGTACAGAAGAAAGGTCGCGATAATTTTGGAAACCA	2812
Db	4231	CAGAAAGTCCCGTAGGCCCAGTCGTAAAGGCTCGGGAATGATAAAGATAAGCCATTGC	4290
QY	2813	CTCCGTTGATGGAAGGTGAGGGGAGATCTTTCAGAGTACTGGGTTCACACCAGAGTCAAA	2872
Db	4291	CTCCTCTGTTGGCCCGTGTGGTGGGAATATTGAAGTACTTGGTTTTTAATGCTCGTCAGC	4350
QY	2873	GGGCCATTTTGTACAGACTTTTGATGAGGTATGG-----AGCTGGCAATTTTGATT	2923
Db	4351	GAAGAAGCTTTCTTAATGCAATTATGCCATATGGTATGCCACCTCAGGATGCTTTTACTA	4410
QY	2924	GGAAGGAGTTTGTTCCTCGCTTAAAGCAGAAGACCTTTGAAGAAATAATGAATATGGAA	2983
Db	4411	CCCAGTGGCTTGTAAAGAGACCTGCGAGGCAAAATCAGAGAAAGAGTTCAAGGCATATGTCT	4470
QY	2984	TACTCTTCTGAAGCACATTGCTGAAGAATAAGACGAGAAATTCCTCAACCTTTTCAGATG	3043
Db	4471	CTCTTTTCATCGGCATTATGTAGCCCGGGGCAGATGGGGCTGAGACCTTTGCTGATG	4530
QY	3044	GTGTGCCCCAAGGAAGCATTAGAAATAGAAGATGTTCTTAGTACAGAAATTCCTCTCTGATAC	3103
Db	4531	GTGTCCCCCGAAGAGGCCTGTCTCGCCAGCATGTCTTACTAGAAATTTGGTGTATGTCTTT	4590
QY	3104	TAGTTTCAGGAGAAGGTGAAATTTGTAGAAAGATCATCCAGGG	3144
Db	4591	TGATTTCGCAAGAAGGTTTCAGGAGTTTGAACATGTTTAATGGG	4631

## RESULT 9

AAD54632  
ID AAD54632 standard; DNA; 6417 BP.

AC AAD54632;

DT 26-JUN-2003 (first entry)

XX DE Human chromodomain helicase DNA binding protein (CHD) encoding DNA #7

Human; p53 pathway; therapeutic; angiogenic disorder; apoptotic disorder;  
KW chromodomain helicase DNA binding protein; CHD; cancer; gene therapy;  
KW cell proliferative disorder; chromatin organisation modifier domain;  
KW cytostatic; ds.

OS Homo sapiens.

PN WQ200298899-A2.

PD 12-DEC-2002.

03-JUN-2002; 2002WO-US017466.

PR 05-JUN-2001; 2001US-0296076P.

PR 22-OCT-2001; 2001US-0338733P.

PR 15-FEB-2002; 2002US-0357600P.

PA (EXEL-) EXELIXIS INC.

PI Friedman L, Plowman GD, Bel

XX  
TJ  
LWJ RRGOTFXX  
XX  
COOZ 'T-TH  
E-

## FRIDELLYNG & CARRO



pt targets for disorders related to defective p53 function e.g. cancer by  
PT contacting an assay system having purified CHD polypeptide or nucleic  
PT acid, with a test agent.

XX  
PS Disclosure; Page 62-66; 278pp; English.

XX  
CC The present invention relates to a method for identifying candidate p53  
CC pathway modulating agents. The method involves contacting an assay system  
CC comprising purified chromatin organisation modifier (chromo) domain  
CC helicase DNA binding proteins (CHD), nucleic acids, their functionally  
CC active fragments or derivatives, with a test agent under conditions  
CC where, but for the presence of the test agent, the system provides a  
CC reference activity. The methods are useful for identifying modulators of  
CC the p53 pathway as therapeutic targets for disorders associated with  
CC defective p53 function, such as angiogenic disorders, apoptotic disorders  
CC or cell proliferative disorders, e.g. cancer. The modulators are useful  
CC as research reagents, diagnostics and therapeutics. The invention is also  
CC useful in gene therapy. The present sequence is human CHD DNA

XX  
SQ Sequence 6417 BP; 1774 A; 1494 C; 1784 G; 1365 T; 0 U; 0 Other;

Query Match 10.1%; Score 421.4; DB 7; Length 6417;  
Best Local Similarity 51.9%; Pred. No. 1.4e-101;  
Matches 1236; Conservative 0; Mismatches 1076; Indels 69; Gaps 10;

QY	818	TACATCCATACCGCTTGAGGGACTTAATTTTTTCGGGTTCTCGTGGTCAAAACAGACGC	877
Db	2266	TGCACCCCTATCAATGGAGGGCCTGAATGGTTGCGCTTCTCCTGGCTCAGGCACTG	2325
QY	878	ATGTAATCCTTGCTGATGAAATGGGACTAGGCAAGACAAATTCAAAGCATTCGCCCTTTAG	937
Db	2326	ACACCATCTTGGCTGATGAGATGGCCCTTGGGAAACTGTACAGACAGCAGTCTTCCTGT	2385
QY	938	CTTCACCTTTTGGAGG-----AACCTCATTCGCGCATTTGGTAATTGCTCCTCTATCGA	991
Db	2386	ATTCCCTTTACAAGGAGGTCATTCCAAAGGCCCTTCTAGTGAGCGGCCCTCTTTCTA	2445
QY	992	CTCTGCGTAACCTGGGAGAGAGAGTTTGGCCACATGGGCCCCACAGATGAACGTGGTTATGT	1051
Db	2446	CCATCATCAACTGGGAGCGGGAGTTTGAAATGTGGGCTCCAGACATGATGTGCTAACT	2505
QY	1052	ATTTTGGCACTGGCGAAGCTCGAGCAGTTATGCGACATGATGAGATGATCAACCTAGATTTCAGCAGTTC	1111
Db	2506	ATGTGGGTGACAAAGGACAGCCGTGCCATCATCCGAGAGAATGAGTTCTCCTTTGAAGACA	2565
QY	1112	ATCAAAAAAAGATCAAGAAAAAGAAATCTGGACAAATAAGTAGCGAAAGCAAGCAAAAAA	1171
Db	2566	ATGCCATTCTGTGTGGCAAGAGGCCTC-----CCGCATGAAGAAAGAGGCAT	2613
QY	1172	GAATCAAGTTTGTATGTCCTCCTCACATCGTATGAGATGATCAACCTAGATTTCAGCAGTTC	1231
Db	2614	CTGTGAAATTCATGTGCTGACATCCTATGAATGATCACCATTGACATGGCTATTT	2673
QY	1232	TAAACCAATTAAGTGGGAGTGCATGATTGTTGATGAAGGTTCATCGACTGAAAAATAAGG	1291
Db	2674	TGGGCTCTATTGATTGGGCTGCCTCATCGTGGATGAAGCCCATCGGCTGAAGAACATC	2733
QY	1292	ATTCAAAGCTGTTCTTCATTGACACAGTATTCAAGTAACCAACCGTATTCTCTGACAG	1351
Db	2734	AGTCTAAGTTCTTCCGGGTATTGAATGGTTACTCACTCCAGCACAAAGCTGTGCTGACTG	2793
QY	1352	GAAACACCACCTTCAGAACAACTTGGATGAACCTTTTCATGCTCATGCAATTTTCTGATCGGG	1411
Db	2794	GGACACCATTACAAAAACAATCTGGAAGAGTTGTTTCATCTGCTCAACTTTCACCCCCCG	2853
QY	1412	GGAAGTTTGAAGTTTGGAGGAGTTCAGGAGGAGTTCAAAGAGTTCAAAGATATTAATCAAGAGGAGC	1471
Db	2854	AGAGGTTCCACAATTTGGAAGGTTTTTTGGAGGAGTTGCTGACATTTGCCAAGGAGGACC	2913
QY	1472	AGATCTCAAGGTTGCACAAAAATGTTGGCTCCACATTTGCTCAGAAGGGTAAAAAAGACG	1531
Db	2914	AGATAAAAAAACTGCATGACATGCTGGGGCCGCACATGTTGGCGCGCTCAAAGCCGATG	2973

QY	1532	TAATGAAAGACATGCCCCCCTCAAAAGGAGCTCATTTTGGTGTGTTGATCTGAGCAGTCTGC	1591
Db	2974	TGTTCAAGAACATGCCCTCCAGACAGAACTAATTGTGCGTGTGGAGCTGAGCCCTATGC	3033
QY	1592	AGAAAGAATATTACAAAGCTATTTTACCCGTAATTATCAAGTATTGA---CAAAAAAGG	1648
Db	3034	AGAAGAAATACTACAAGTACATCCTCACTCGAAATTTTGAAGCACTCAATGCCGAGGTG	3093
QY	1649	GAGGTGCTCAAAATTTCCCTTAATAACATTTATGATGGAATACGAAAAAGTATGCTGCCATC	1708
Db	3094	GTGGCAACCAGGTGCTCTCTGCTGAATGTGGTATGATGATCTTAAGAAGTGTGCAACCATC	3153
QY	1709	CTTA-----TATGCTAGAGGGTGTGAGCCAGTTTATTCACGACGCAAAATGAAGCTT	1759
Db	3154	CATACCTCTTCCCTGTGGCTGCAATGGAAGCTCCTAAGATGCCTAATGGCATGTATGATG	3213
QY	1760	TCAACAACATTTTGGAGTCTTGTGGAAGCTGCAACTTCTAGATAAAATGATGTTCAAAC	1819
Db	3214	GCAGTCCCTAATCAGAGCATCTGGGAAATATTATGCTGCTGCAGAAAAATGCTCAAGAACC	3273
QY	1820	TGAAGAGCAAGGACACAGAGTCTCTAATAATATACACACAGTTTCAGCATATGCTGGACTTAC	1879
Db	3274	TTAAGGAGGTGGGCATCGTGTACTCATCTTTTCCCAGATGACCAAGATGCTAGACCTGC	3333
QY	1880	TTGAAGACTACTGTACCCATAAGAAATGGCAGTACGAGCGCAATGATGGAAGGTTGGCG	1939
Db	3334	TAGAGGATTTCTTGGAACATGAAGTTATAAATACGAACGCATCGATGGTGAATCACTG	3393
QY	1940	GAGCTGACGGCAAAATACGCATAGATCGGTTCAATGCCAAAAATTTAAACAAGTTTGTGT	1999
Db	3394	GGAACATGGGCAAGAGGCCATTGACCGCTTCAATGCACCCGGTGTCTCAGCAGTTCTGCT	3453
QY	2000	TTTTGCTCTCCACAAGAGCTGGTGGCTTAGGAATAAATCTTGCACCGCTGATACAGTAA	2059
Db	3454	TCCTTGCTTTCCACTCGAGCTGGGGCCTTGGAATCAATCTGGCCACTGCTGACACAGTTA	3513
QY	2060	TCATTTATGACAGTGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGTCAATC	2119
Db	3514	TTATCTATGACTCTGACTGGAAACCCCATTAATGACATTCAGGCCCTTTAGCAGAGTCAAC	3573
QY	2120	GACTTGGCCAAAACAAATAAGGTGATGATTTATAGGCTCATAAACGAGGCACCATTTGAAG	2179
Db	3574	GGATTGGGCAAAAATAAAAAGGTAATGATCTACCGGTTTGTGACCGCTGCTCAGTGGAGG	3633
QY	2180	AAAGGATGATCAATTGACTAAAAAGAAAAATGGTTCTAGAGCATCTTGTGTGG-----	2234
Db	3634	AGCGCATCAGCAGGTGGCAAGAAAGAAAAATGATGCTGACGCATCTAGTGTGCGGCCTG	3693
QY	2235	---GAACTCAAAACACAAAACATTAATCAGGAAGAGTTAGATGACATCATCAGGTATG	2290
Db	3694	GGCTGGGCTCCAAGACTGGATCTATGTCCAAACAGGAGCTTGTATGATATCCTCAAATTTG	3753
QY	2291	GATCAAAGGAGCTTTTGTCTAGTGAAGATGAAGCAGGAAAGTCTGGAAAAAATTCATT	2350
Db	3754	GCACAGGAACTATTCAAGGATGAAGCCACTGATGGAGGAGGAGACAAACAAAGAGGGAG	3813
QY	2351	ATGATGATGCGGCTATAGACAAATTCCTTGA-----TCGTGATCTCGTGGAGGCAGAGG	2404
Db	3814	AAGATAGCAGTGTATCCACTACGATGATAAGGCCATTGAACCGCTGCTAGACCGTAACC	3873
QY	2405	AAGTCTCAGTGGATGAAGAGGAGATGGATTCTTAAAGGCTTTCAAGGTGGCTAATT	2464
Db	3874	AGGATGAGACTGAAGACACAGAAATTCAGGGGCATGAATGAATATTTGAGCTCATTCAAAG	3933
QY	2465	TTGAATATATAGATGAAAAATGAGGCAGCAGCATTAGAGGCACAGAGAGTCGCTGCTGAAA	2524
Db	3934	TGGCCAGTATGTGTACGGGAAGAAATGGGGAGGAAGAGGAGGTAGAACGGGAAA	3993
QY	2525	GCAATCTTCAGCAGGCAATTTCTGATAGACCAAGTTATTGGGAAGAGTTGTTAAAGATA	2584
Db	3994	TCATTTAAACAGGAAGAAAGTGTGGATCCTGA---CTACTGGGAGAAATGCTGCGGCACC	4050
QY	2585	AATTGAGCTGCACCCAGGCTGAGGAGCTTAATGCTCTTTGGAAAAAAGGAAGAGATCGCA	2644

Db 4051 ATTATGAGCAGCAGCAAGAAGATCTAGCCCGAAATCTGGGCAAAAGGAAAAAGATCCGTA 4110  
QY 2645 AGCAGTTGGTATCCATTGAAG--AAGATGATCTTGCTGGTTTGGAGATGTGAGCTCTG 2701  
Db 4111 AACAGGTCAACTACAATGATGGCTCCAGGAGCGAGATTGGCAGACGACCACTCCG 4170  
QY 2702 ATGGAGATGAAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGACAAAGGATT 2761  
Db 4171 ACAACCAAGTCCGATTACTCAGTGGCTTCAGAGGAAGGTGATGAAGACTTTGATGAACGTT 4230  
QY 2762 AGACGGGTGACGCGCGTACAGAAAGAGGTCGCGATAATTTGGAAACAA----- 2812  
Db 4231 CAGAAAGTCCCCGAGGCCCAAGTCGTAAGGGCTCGGAATGATAAAGATAAGCCATTGC 4290  
QY 2813 CTCGGTTGATGGAAGGTGAGGGGAGATCTTTACAGACTACTGGTTTCAACCAAGATCAAA 2872  
Db 4291 CTCCTCTGTTGGCCCGTGTGGTGGGAATATGAAGTACTTGGTTTAAATGCTCGTCAGC 4350  
QY 2873 GGGCCATTTTGTACAGACTTTGTATGAGGTATGG-----AGCTGGCAATTTTGATT 2923  
Db 4351 GAAAAGCCTTTCTTAATGCAATTATGCGATATGGTATGCCACCTCAGGATGCTTTTACTA 4410  
QY 2924 GGAAGGAGTTTGTCTCTCGCTTAAAGCAGAGACCTTTTGAAGAAATTAATGAATATGGAA 2983  
Db 4411 CCCAGTGGCTTTGAAGAGACCTGCGAGGCAATCAGAGAAAGAGTTCAAGGCATATGTCT 4470  
QY 2984 TACTCTCTTGAAGCACATTTGCTGAAGAAATAGACGAGAATTTCTCCAACCTTTTCAGATG 3043  
Db 4471 CTCCTTTTCATCGGCATTATGTGAGCCCGGGGCAGATGGGGCTGAGACCTTTGCTGATG 4530  
QY 3044 GTGTGCCCCAAGGAGGACTTAGAATAGAAGATGTTCTAGTCAGAAATGCTTCTTGATAC 3103  
Db 4531 GTGTCCCCCGAGAAGGCCCTGTCTCGCCAGCATGTCTTACTAGAAATGGTGTATGTCTT 4590  
QY 3104 TAGTTCAGGAGAAAGTGAAATTTGTAGAAGATCATCCAGGG 3144  
Db 4591 TGATTGCAAGAAAGGTTCAGGAGTTTGAACATGTTTAATGGG 4631

RESULT 10

AAI58444  
ID AAI58444 standard; cDNA; 6475 BP.  
XX  
AC AAI58444;  
XX  
DT 22-OCT-2001 (first entry)  
XX  
DE Human polynucleotide SEQ ID NO 647.  
XX  
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia; ss.  
OS Homo sapiens.  
XX  
PN WO200153312-A1.  
XX  
PD 26-JUL-2001.  
XX  
PF 26-DEC-2000; 2000WO-US034263.  
XX  
PR 23-DEC-1999; 99US-00471275.  
PR 21-JAN-2000; 2000US-00488725.  
PR 25-APR-2000; 2000US-00552317.  
PR 20-JUN-2000; 2000US-00598042.  
PR 19-JUL-2000; 2000US-00620312.  
PR 03-AUG-2000; 2000US-00653450.  
PR 14-SEP-2000; 2000US-00662191.  
PR 19-OCT-2000; 2000US-00693036.

PR 29-NOV-2000; 2000US-00727344.  
XX (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
PI Zhou P, Goodrich R, Drmanac RT;  
XX  
DR WPI; 2001-442253/47.  
DR P-PSDB; AAM39288.  
XX  
PT Novel nucleic acids and polypeptides, useful for treating disorders such  
PT as central nervous system injuries.  
XX  
PS Claim 1; SEQ ID NO 647; 10078pp; English.  
XX  
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the  
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,  
CC immunosuppressant and cyostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders. Note: The sequence data for this patent did not form  
CC part of the printed specification  
XX  
SQ Sequence 6475 BP; 1760 A; 1528 C; 1822 G; 1365 T; 0 U; 0 Other;  
Query Match 10.1%; Score 421.4; DB 4; Length 6475;  
Best Local Similarity 51.9%; Pred. No. 1.4e-101;  
Matches 1236; Conservative 0; Mismatches 1076; Indels 69; Gaps 10;  
QY 818 TACATCCATACCAGCTTGAGGGACTTAATTTTTTTCGGTTCTCGTGGTCAAAACAGACGC 877  
Db 2395 TGCACCCCTATCAATGGAGGGCCTGAATGGTTGCGCTTCTCTGGGCTCAGGGCACTG 2454  
QY 878 ATGTAATCCTTGCTGATGAAATGGGACTAGGCAAGACAATTCAAAGCATTGCCCTTTTAG 937  
Db 2455 ACACCATCTTGGCTGATGAGATGGGCCCTTGGGAAAACCTGTACAGACAGCAGTCTTCTGT 2514  
QY 938 CTTCACTTTTGGAGGAG-----AACCTCATTCGCAATTTGGTAATTGCTCCTCTATCGA 991  
Db 2515 ATTCCCTTTACAAGGAGGGTCAATTCAAAGGCCCTTCTCCTAGTAGCGGCCCTCTTTCTA 2574  
QY 992 CTCTGCGTAACCTGGGAGAGAGAGTTTGCCACATGGGCCCCACAGATGAACGTGGTTATGT 1051  
Db 2575 CCATCATCAACTGGGAGCGGGAGTTGAAATGTGGCTCCAGACATGTATGTCGTAACCT 2634  
QY 1052 ATTTTGGCACTGCGCAAGCTCGAGCAGTTATCAGAGAACATGAGTTTACTTATCGAAAG 1111  
Db 2635 ATGTGGGTGACAAGGACAGCCCGTGCCATCATCCGAGAGATGAGTTCTCCTTTGAAGACA 2694  
QY 1112 ATCAAAAAAAGATCAAGAAAAAGAAATCTGGACAAATAAGTAGCGAAAGCAAGCAAAAA 1171  
Db 2695 ATGCCATTCTGTTGGTGGCAAGAGGCCCTC-----CCGATGAAGAAAGAGGCAT 2742  
QY 1172 GAATCAAGTTTGATGTCTCTCTCACATCGTATGAGATGATCAACCTAGATTTCAGCAGTTC 1231  
Db 2743 CTGTGAAATTCATGTCTGTCTGACATCCTATGAATGATCAACCATGACATGGCTATTT 2802  
QY 1232 TAAACCAATTAAGTGGGAGTGCATGATTTGTTGATGAAGGTGCATCGACTGAAAAATAAGG 1291  
Db 2803 TGGGCTCTATTGATTGGGCCCTGCCTCATCGTGGATGAAGCCCATCGGCTGAAGAACAAATC 2862  
QY 1292 ATTCAAAGCTGTTCTCTTCATTGACACAGTATTCAAGTAACCCGTTATTTCTTGACAG 1351  
Db 2863 AGTCTAAGTTCTTCCGGGTATTGAATGGTTACTCACTCCAGCACAAGCTGTTGCTGACTG 2922

QY	1352	GAACACCACCTTCAGAACAACTTGGATGAACCTTTTCATGCTCATGCAATTTTCTTGATGCGG	1411
Db	2923	GGACACCATTACAAAACAATCTGGAAGAGTTGTTTCATCTGCTCAACTTCTCACCCCG	2982
QY	1412	GGAAGTTTGGAGTTTGGAGGAGTTCAGGAGGAGTTCAAAGATATTTAATCAAGAGGAGC	1471
Db	2983	AGAGGTTCCACAATTGGAAAGGTTTTTTGGAGGAGTTTGCTGACATTGCCAAGGAGGACC	3042
QY	1472	AGATCTCAAGGTTGCACAAAAATGTTGGCTCCACATTTGCTCAGAAAGGTAAAAAAAGACG	1531
Db	3043	AGATAAAAAAACTGCATGACATGCTGGGCCGCACATGTTGCGCGGCTCAAAGCCGATG	3102
QY	1532	TAATGAAAGACATGCCCCCAAAAAAGGAGCTCAATTTGCGTGTGATCTGAGCAGTCTGC	1591
Db	3103	TGTTCAAGAACATGCCCTCCAAGACAGAACTAATTGTGCGTGTGGAGCTGAGCCCTATGC	3162
QY	1592	AGAAAGAATATTACAAAGCTATTTTACCCGTAATTATCAAGTATTGA--CAAAAAAGG	1648
Db	3163	AGAAGAAATACTACAAGTACATCCTCACTCGAAATTTTGAAGCACTCAATGCCGAGGTG	3222
QY	1649	GAGGTGCTCAAAATTTCCCTTAATAACATTATGATGGAATTACGAAAAGTATGCTGCCATC	1708
Db	3223	GTGGCAACCAGGTGTCGTCTGCTGTAATGTGGTGATGGATCTTAAGAAGTCTGCAACCATC	3282
QY	1709	CTTA-----TATGCTAGAGGGTGTGAGCCAGTTATTACGACGCAAAATGAAGCTT	1759
Db	3283	CATACCTCTTCCCTGTGGCTGCAATGGAAGCTCCTAAGATGCCTAATGGCATGTATGATG	3342
QY	1760	TCAAAACAACCTTTTGGAGTCTTGTGGAAAGCTGCAACTTCTAGATAAAATGATGGTCAAAC	1819
Db	3343	GCAGTGCCCTAATCAGAGCATCTGGGAAATTATTGCTGCTGCAGAAAAATGCTCAAGAAC	3402
QY	1820	TGAAAGAGCAAGGACACAGAGTCCTAATATACACACAGTTTCAGCATATGCTGGACTTAC	1879
Db	3403	TTAAGGAGGGTGGGCATCGTGACTCATCTTTTCCAGATGACCAAGATGCTAGACCTGC	3462
QY	1880	TTGAAGACTACTGTACCCATAAGAAATGGCAGTACGAGCGAAATTGATGGAAAGGTTGGCG	1939
Db	3463	TAGAGGATTTCTTGGAAACATGAAGGTTATAAATACGAACGCATCGATGGTGGAAATCA	3522
QY	1940	GAGCTGAGCGGCAATACGCATAGATCGGTTCAATGCCAAAAAATTTCTAACAAAGTTTGT	1999
Db	3523	GGAACATGCGGCAAGAGGCCATTGACCGCTTCAATGACCCGGTGTCTCAGCAGTTCTGCT	3582
QY	2000	TTTTGCTCTCCACAAGAGCTGGTGGCTTAGGAATAAAATCTTGCAACGGCTGATACAGTAA	2059
Db	3583	TCTTGCTTTCCACTCGAGCTGGGGCCCTTGGAAATCAATCTGGCCACTGTGACACAGTTA	3642
QY	2060	TCATTTATGACAGTGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC	2119
Db	3643	TTATCTATGACTCTGACTGGAACCCCAATAATGACATTCAGGCCCTTTAGCAGAGCTCACC	3702
QY	2120	GACTTGGCCAAACAATAAGGTGATGATTTATAGGCTCATAAACCGAGGCAACCATTGAAG	2179
Db	3703	GGATTGGGCAAAATAAAAGGTAATGATCTACCGGTTTGTACCCGTGCGTCAGTGGAGG	3762
QY	2180	AAAGGATGATGCAATTGACTAAAAAGAAAATGGTTCTAGAGCATCTTGTGTTGG-----	2234
Db	3763	AGCGCATCAGCAGGTGGCAAAAGAAAATGATGCTGACGCATCTAGTGGTGGGCCTG	3822
QY	2235	----GAAACTCAAAACAAAAACATTAAATCAGGAAGAGTTAGATGACATCATCAGGTATG	2290
Db	3823	GGCTGGGCTCCAAGACTGGATCTATGTCCAACAGGAGCTTGTATGATATCCTCAAAATTG	3882
QY	2291	GATCAAAGGAGCTTTTGTGCTAGTGAAGATGATGAAGCAGGAAAAGTCTGGAATAATTCATT	2350
Db	3883	GCACTGAGGAACATTCAAGGATGAAGCCACTGATGGAGGAGGAGACAAACAAGAGGGAG	3942
QY	2351	ATGATGATCGGGCTATAGACAAATTTGCTTGA-----TCGTGATCTCGTGGAGGCAGAGG	2404
Db	3943	AAGATAGCAGTGTATTCCACTACGATGATGAAGGCCATTGAACGGCTGCTAGACCGTAACC	4002

QY	2405	AAGTCTCAGTGGATGATGAAGAGGAGAAATGGATTCTTAAAGGCTTTCAGGTGGCTAATT	2464
Db	4003	AGGATGAGACTGAAGACACAGAAATTGCAGGGCATGAATGAATATTGAGCTCATTTCAAAG	4062
QY	2465	TTGAATATATAGATGAAAAATGAGGCAGCAGCATTAGAGGCACAGAGAGTCGTGCTGAAA	2524
Db	4063	TGGCCCACTATGTGTACGGGAAGAAGAAATGGGGAGGAAGAGGAGGTAGAACGGGAAA	4122
QY	2525	GCAAAATCTTCAGCAGGCAATTCGTATAGAGCAAGTTATTGGGAAGAGTTGTTAAAAAGATA	2584
Db	4123	TCATTAAACAGGAAGAAAGTGTGGATCCTGA---CTACTGGGAGAAATTGTGCGGCACC	4179
QY	2585	AATTTGAGCTGCACCAGGCTGAGGAGCTTAATGCTCTTGGAAAAAGGAAGAGTCGCA	2644
Db	4180	ATTATGAGCAGCAGCAAGAAGATCTAGCCCCGAAATCTGGGCAAGGAAAAAGAAATCCGTA	4239
QY	2645	AGCAGTTGGTATCCATTGAAG---AAGATGATCTTTGCTGGTTTGGAAAGATGTGAGCTCTG	2701
Db	4240	AACAGGTCAACTACAATGATGGTCCCAGGAGGACCGAGATTGGCAGGACGACCAGTCCG	4299
QY	2702	ATGGAGATGAAAAGTTTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGACAAGGAGTTC	2761
Db	4300	ACAACCAGTCCGATTACTCAGTGGCTTCAGAGGAAGGTGATGAAGACTTTGATGAACGTT	4359
QY	2762	AGACGGTCCGACGGCCGTACAGAAAAAGGGTCGCGATAATTTTGGAAACCAA-----	2812
Db	4360	CAGAAGCTCCCCGTAGGCCCAGTCGTAAGGCCCTGCGGAATGATAAAAGATAAGCCATTGC	4419
QY	2813	CTCCGTTGATGGAAGGTGAGGGGAGATCTTTTCAGAGTACTGGGTTTCAACCAGAGTCAAA	2872
Db	4420	CTCCTCTGTTGGCCCGTGTGTTGGTGGGAATATTGAAGTACTTGGTTTAAATGCTCGTCAGC	4479
QY	2873	GGGCCATTTTGTACAGACTTTTGATGAGGTATGG-----AGCTGGCAATTTTGATT	2923
Db	4480	GAAAAGCCTTCTTAATGCAATTATGCGATATGGTATGCCACCTCAGGATGCTTTTACTA	4539
QY	2924	GGAAGGAGTTTGTTCCTCGCTTAAAGCAGAAGACCTTTTGAAGAAATAAATGAATATGGAA	2983
Db	4540	CCCAGTGGCTTGTAAAGAGACCCTGCGAGGCAAAATCAGAGAAAGAGTTCAAGGCATATGTCT	4599
QY	2984	TACTCTTCTTGAAGCACATTGCTGAAGAAAAATAGACGAGAAATTCCTCAAACCTTTTCAGATG	3043
Db	4600	CTCTTTTCATGCGGCATTTATGTGAGCCGGGCGAGATGGGCTGAGACCTTTTGCTGATG	4659
QY	3044	GTGTGCCCAAGGAAGGACTTAGAATAGAAGATGTTCTAGTCAGAATTGCTCTTCTGATAC	3103
Db	4660	GTGTCCCCCGAGAAGGCCCTGTCTCGCCAGCATGTCCTTACTAGAAATTGGTGTATGTCTT	4719
QY	3104	TAGTTCAGGAGAAGGTGAAATTTGTAGAAGATCATCCAGGG	3144
Db	4720	TGATTCCGAAGAAGGTTTCAGGAGTTTGAACATGTTAATGGG	4760
RESULT 11			
ADB48415			
ID	ADB48415	standard; cDNA; 6475 BP.	
XX	AC	ADB48415;	
XX	DT	04-DEC-2003 (first entry)	
XX	DE	Novel human cDNA SEQ ID NO 325.	
XX	KW	ss; cancer; neurodegenerative disease; human.	
XX	OS	Homo sapiens.	
XX	PN	US2003104529-A1.	
XX	PD	05-JUN-2003.	
XX	PF	04-JAN-2002; 2002US-00037270.	
XX			



PR 21-JAN-2000; 2000US-00488725.  
PR 25-APR-2000; 2000US-00552317.  
PR 19-JUL-2000; 2000US-00620312.  
XX  
PA (ZHOU/) ZHOU P.  
PA (TANG/) TANG Y T.  
PA (LIUC/) LIU C.  
PA (ASUN/) ASUNDI V.  
PA (DRMA/) DRMANAC R T.  
XX  
PI Zhou P, Tang YT, Liu C, Asundi V, Drmanac RT;  
XX WPI; 2003-678194/64.  
DR  
XX New polynucleotide, useful for treating diseases e.g., cancer or  
PT neurodegenerative diseases.  
XX  
PS Claim 1; SEQ ID NO 325; 99pp; English.  
XX  
CC The invention relates to a polynucleotide comprising a sequence given in  
CC the specification, or its mature protein-coding portion, or its  
CC complement. The polynucleotide is useful for treating diseases e.g.,  
CC cancer or neurodegenerative diseases and many others listed in the  
CC specification. The present sequence represents a novel human cDNA. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format directly from USPTO  
CC at seqdata.uspto.gov/sequence.html?DocID=20030104529.  
XX  
SQ Sequence 6475 BP; 1760 A; 1528 C; 1822 G; 1365 T; 0 U; 0 Other;  
  
Query Match 10.1%; Score 421.4; DB 8; Length 6475;  
Best Local Similarity 51.9%; Pred. No. 1.4e-101;  
Matches 1236; Conservative 0; Mismatches 1076; Indels 69; Gaps 10;  
  
QY 818 TACATCCATACCAGCTTGAGGGACTTAATTTTTCGCGTTCTCGTGGTCAAAAACAGACGC 877  
DB 2395 TGCACCCCTATCAAAATGGAGGGCCTGAATTGTTGCGCTTCTCCTGGGCTCAGGGCACTG 2454  
  
QY 878 ATGTAATCCTTGCTGATGAAATGGGACTAGSCAAGACAATTCAAAGCATTTGCCCTTTTAG 937  
DB 2455 ACACCATCTTGGCTGATGAGATGGGCCCTTGGGAAAACGTGTACAGACAGCAGTCTTCCTGT 2514  
  
QY 938 CTTCACTTTTGGAGG-----AACCTCATTCGGCATTTGGTAATTGCTCCTCTATCGA 991  
DB 2515 ATTCCCTTTACAGGAGGGTCAATCCAAAGGCCCTTCTAGTAGCGGCCCTCTTTCTA 2574  
  
QY 992 CTCTGCGTAACCTGGGAGAGAGAGTTTGGCCACATGGGCCCCACAGATGAACGTGGTTATGT 1051  
DB 2575 CCATCATCAACTGGGAGCGGAGTTTGAAATGTGGGCTCCAGACATGTATGTCGTAACCT 2634  
  
QY 1052 ATTTTGGCACTGCGCAAGCTCGAGCAGTTATCAGAGAACATGAGTTTACTTATCGAAAG 1111  
DB 2635 ATGTGGGTGACAAGGACAGCCGTGCCATCATCCGAGAGAATGAGTTCTCCTTTGAAGACA 2694  
  
QY 1112 ATCAAAAAAAGATCAAGAAAAAGAAATCTGGACAAATAAGTAGCGAAAGCAAGCAAAAA 1171  
DB 2695 ATGCCATTTCGTGGTGGCAAGAGGCCCTC-----CCGCATGAAGAAAGAGGCAT 2742  
  
QY 1172 GAATCAAGTTTGATGCTCCTCCATCATCGTATGAGATGATCAACCTAGATTTCAGCAGTTC 1231  
DB 2743 CTGTGAAATTCCATGTGCTGTGACATCCATGAATTGATCACCATTGACATGGCTATTT 2802  
  
QY 1232 TAAACCAATTAAGTGGGAGTGCATGATTGTTGATGAAGGTCAATCGACTGAAAAATAAGG 1291  
DB 2803 TGGGCTCTATTGATTGGGCTGCCTCATCGTGGATGAAGCCCATCGGCTGAAGAAACAATC 2862  
  
QY 1292 ATTCAAAGCTGTTCTCTTCATTGACACAGTATTCAAGTAACCAACCGTATTCTTCTGACAG 1351  
DB 2863 AGTCTAAGTTCTTCGGGTATTGAATGGTTACTCACTCCAGCACAAGCTGTGCTGACTG 2922  
  
QY 1352 GAACACCACCTTCAGAACAACTTGGATGAACCTTTTCATGCTCATGCAATTTCTTGATCGGG 1411  
DB 2923 GGACACCATTACAAAAACAATCTGGAAGAGTTGTTTCATCTGCTCAACTTCTCAACCCCG 2982

QY 1412 GGAAGTTTGGAAAGTTTGGAGGAGTTCCAGGAGGAGTTCAAAGATATTAATCAAGAGGAGC 1471  
DB 2983 AGAGGTTCCACAATTTGGAAGGTTTGGAGGAGTTTGGAGGAGTTTGGAGGAGGACC 3042  
  
QY 1472 AGATCTCAAGGTTGCACAAAATGTGGCTCCACATTTTGTCTAGAAAGGTTAAAAAAGACG 1531  
DB 3043 AGATAAAAAAACTGCATGACATGTCTGGGCGCCGACATGTTGCGCGGCTCAAAGCCGATG 3102  
  
QY 1532 TAATGAAGACATGCCCCCCCCAAAAGGAGCTCATTTTGGTGTGTGATCTGAGCAGTCTGC 1591  
DB 3103 TGTTCAGAACAATGCCCTCCAAAGACAGAACTAATTTGCGTGTGAGCTGAGCCCTATGC 3162  
  
QY 1592 AGAAAGAATATTACAAAGCTATTTTACCCGTAATTAATCAAGTATTGA---CAAAAAAGG 1648  
DB 3163 AGAAGAAATACTACAAGTACATCCTCACTCGAAATTTTGAAGCACTCAATGCCGAGGTG 3222  
  
QY 1649 GAGGTGCTCAAATTTCCCTTAATAACATTAATGATGGAATTACGAAAAAGTATGCTGCCATC 1708  
DB 3223 GTGGCAACCAGGTGTCTCTGCTGAATGTGGTATGATGATCTTAAGAAGTGTGCAACCATC 3282  
  
QY 1709 CTTA-----TATGCTAGAGGGTGTGAGCCAGTTATTACGACGCAAAATGAAGCTT 1759  
DB 3283 CATACCTCTTCCCTGTGGCTGCAATGGAAGCTCCTAAGATGCCTAATGATGATGATG 3342  
  
QY 1760 TCAAAACAACCTTTTGGAGTCTTGTGGAAGCTGCAACTTCTAGATAAAATGATGTCAAAC 1819  
DB 3343 GCAGTGCCCTAATCAGAGCATCTGGGAAATATTGCTGCTGCAGAAAAATGCTCAAGAACC 3402  
  
QY 1820 TGAAGAGCAAGGACACAGAGTCCTTAATATATACACACAGTTTTCAGCATATGCTGGACTAC 1879  
DB 3403 TTAAGGAGGGTGGGCATCGTGTACTCATCTTTCCAGATGACCAAGATGCTAGACCTGC 3462  
  
QY 1880 TTGAAGACTACTGTACCCCATAGAATGGCAGTACGAGCGCAATTGATGGAAGGTTGGCG 1939  
DB 3463 TAGAGGATTTCTTGGAAACATGAAGTTATAAATCGAACGCTGATGTTGGAATCACTG 3522  
  
QY 1940 GAGCTGAGCGGCAAAATACGCATAGATCGGTTCAATGCCAAAAATTTCTAACAAAGTTTGT 1999  
DB 3523 GGAACATGCGGCAAGAGGCCATTGACCGCTTCAATGCACCGGCTGCTCAGCAGTTCTGCT 3582  
  
QY 2000 TTTTGTCTCCACAAGAGCTGTGGCTTAGGAATAAATCTTGCAACGGCTGATACAGTAA 2059  
DB 3583 TCTTGTCTTCCACTCGAGCTGGGGCTTGAATCAATCTGGCACTGCTGACACAGTTA 3642  
  
QY 2060 TCATTTATGACAGTGACTGGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC 2119  
DB 3643 TTATCTATGACTCTGACTGGAACCCCATTAATGACATTCAGSCCTTTAGCAGAGCTCAC 3702  
  
QY 2120 GACTTGGCCAAACAAATAAGGTGATGATTTATAGGCTCATAAACCGAGCACCATTTGAAG 2179  
DB 3703 GGATTGGCAAAATAAAAAGGTAATGATCTACCGGTTTGTGACCGCTGCTCAGTGGAGG 3762  
  
QY 2180 AAAGGATGATGCAATTGACTAAAAGAAAATGGTTCTAGAGCATCTTGTGTTGG----- 2234  
DB 3763 AGCGCATCAGCAGGTGGCAAGAAAGAAAATGATGCTGACGCACTAGTGGTGGCGCCTG 3822  
  
QY 2235 ----GAAACTCAAAACACAAAACATTAATCAGGAAGAGTTAGATGACATCATCAGGTATG 2290  
DB 3823 GGCTGGGCTCCAAGACTGGATCTATGTCCAACACAGGAGCTTGTATGATATCTCAAATTTG 3882  
  
QY 2291 GATCAAGGAGCTTTTGTCTAGTGAAGATGATGAAGCAGGAAAGTCTTGAAAAATTCATT 2350  
DB 3883 GCACCTGAGGAACATTTCAAGGATGAAGCCACTGATGGAGGAGAGACAAACAAGAGGGAG 3942  
  
QY 2351 ATGATGATCGGGCTATAGACAAATTGCTTGA-----TCGTGATCTCGTGGAGGCAGAGG 2404  
DB 3943 AAGATAGCAGTGTATTCCACTACGATGATAAGGCCATTGAACGGCTGTAGACCGTAACC 4002  
  
QY 2405 AAGTCTCAGTGGATGATGAAGAGAGAATGGATTCTTAAAGGCTTTCAAGGTGGCTAATT 2464  
DB 4003 AGGATGAGACTGAAGACACAGAAATTCAGGGGCATGAATGAATATTGAGCTCATTCAAAG 4062



Db 2937 GCTGACAGGAACCCCATTCAGAAATAATCTGGAGAGCTCTTCCATCTCCTGAACTTCCT 2996

Qy 1404 TGATCGGGGAAGTTTGGAAAGTTTGGAGGAGTTCAGGAGGAGTTCAAAGATATTAATCA 1463

Db 2997 CACCCAGAGAGATTAAACAACCTTGGAGGGCTTCTGGAGGAGTTTGTGACATATCCAA 3056

Qy 1464 AGAGGAGCAGATCTCAAGGTTGCACAAAATGTGGCTCCACATTTTGTCTCAGAAGGGTAAA 1523

Db 3057 AGAGGACCAGATCAAGAAACTGCATGATTTGCTGGGGCCACACATGTCTGGGAGACTCAA 3116

Qy 1524 AAAAGACGTAATGAAAGACATGCCCCCCCCAAAAGGAGCTCATTTTGGGTGTTGATCTGAG 1583

Db 3117 GGCAGATGCTTTAAGAACATGCCAGCCAAAGACAGAGCTCATCGTTCCGGTGGAGCTAAG 3176

Qy 1584 CAGTCTGCAGAAAGAAATATTACAAAGCTAATTTTACCCTAATTAATCAAGTATTGA ---C 1640

Db 3177 CCCCATGCAGAAAGAAATACTACAAATACATCCTGACTCGAAATTTTGAGGCCCTTGAATTC 3236

Qy 1641 AAAAAAGGAGGTGCTCAAAATTTCCCTTAATAAACATTAATGATGGAATTACGAAAAGTATG 1700

Db 3237 ACGAGGTGTTGGGAACCAAGGTGTCGCTGCTTAATATCATGATGATCTTAAGAAGTGTG 3296

Qy 1701 CTGCCATCCTTATATCTAGAGGGTGTG-----AGCCAGTTATTACGACGCAAA 1751

Db 3297 CAACCATCCATACCTTTTCCCGTGGCTGATGGAGTCCCCCAAATCTCCCGAGTGGGC 3356

Qy 1752 TGAAGCTTTCAAACAACTTTTGGAGCTTTTGGAAAGCTGCAACTTCTAGATAAAATGAT 1811

Db 3357 TTATAGGGTGGGCACCTTATTAAAGTCGTGCGGAAGCTCATGCTGCCAGAAAGATGCT 3416

Qy 1812 GGTCAAACTGAAAGCAAGGACACAGAGTCTCTAATAATACACACAGTTTCAGCATATGCT 1871

Db 3417 GCGAAAGCTGAAGGAGCAAGGACACCGAGTGTCTCATCTTCTCGCAGATGACCAAAATGTT 3476

Qy 1872 GGACTTACTTGAAGACTACTGTACCCCATAGAAATGGCAGTACGAGCGAATTGATGAAA 1931

Db 3477 AGACTTGCTTGAGACTTCTTAGACTATGAAGGCTACAAGTATGAGCGCATCGATGGTGG 3536

Qy 1932 GGTGGCGGAGCTGAGCGGCAAAATACGCATAGATCGGTTCAATGCCAAAAATCTTAACAA 1991

Db 3537 TATCACGGGTGCCCTGAGGCAGGAGGCCATCGATCGGTTAATGCTCCTGGGGCCCCAACA 3596

Qy 1992 GTTTTGTGTTTTGCTCTCCACAAGAGCTGGTGGCTTAGGAATAAATCTTGCAACGGCTGA 2051

Db 3597 ATTCTGCTTCTCTGTCCACCCGAGCTGGGGCCCTGGGCATCAATCTGCCACATGCTGA 3656

Qy 2052 TACAGTAATCATTTATGACAGTGAATCGCTCATGCTGATCTTCAAGCAATGGCTAG 2111

Db 3657 CACTGTATCATCTTTGATTTCTGACTGGAAACCCCATTAATGACATCCAGGCCCTTAGCCG 3716

Qy 2112 AGCTCATCGACTTGGCCAAACAAATAAGGTGATGATTTATAGGCTCATAAACCGAGGCAC 2171

Db 3717 GGCTCATCGGATTGGCCAGGCCAACAAAGTATGATGATTACCGGTTTGTGACTCGCGGTC 3776

Qy 2172 CATTTGAAGAAAGGATGATGCAATTGACTAAAAAGAAAATGGTTCTAGAGCATCTTGTGTT 2231

Db 3777 AGTGAAGAGCGAATCACAAAGTGGCCCAAGAGAAAGATGATGCTGACACACACCTGTTGT 3836

Qy 2232 TGG-----GAAACTCAAAACACAAACATTAATCAGGAAGAGTTAGATGACATCAT 2282

Db 3837 GCGGCTGGGCTGGGCTCCAAGGCAGGCTCCATGTCCAAGCAGGAGCTTGACGACATTCT 3896

Qy 2283 CAGGTATGGATCAAAGGAGCTTTTGTAGTGAAGATGATGAAGCAGGAAAGTCTG----- 2338

Db 3897 CAAATTTGGCACTGAAGAGCTATTCAAGGATGAAAACGAGGGGAGAAACAAGGAGGAGGA 3956

Qy 2339 -----GAAAAATTCATTTATGATGATGCGGCTATAGACAAATTTGCTTGATCGTATCTCGT 2393

Db 3957 CAGCAGTGTGATTTCATTATGACAAATGAGGCCATCGCTCGGCTGTTGGACCGGAACAGGA 4016

Qy 2394 GGAGGCAGAGGAAGTCTCAGTGGATGATGAAGAGGAGAAATGGATTCTTAAAGGCTTTCAA 2453

Db 4017 TGCAACTGAGGA-----CACTGACGTGCAGAAACATGAATGATATCTCAGCTCCTTCAA 4070

Qy 2454 GGTGGCTAATTTTGAATATATAGATGAAAAATGAGGCAGCAGCATTAGAGGCACAGAGAT 2513

Db 4071 GGTGGCAGTACGTCTGTCGGGAAGAAGACAAGATTGAGGAATTTAGCGGAGATCAT 4130

Qy 2514 CGTGTCTGAAAAGCAAATCTTACGAGGCAATTTCTGATAGACGAAGTTATTGGGAAGATT 2573

Db 4131 CAAGCAGGAGGAGAAATGTGGACCTTGAC-----TACTGGGAGAAAGCT 4172

Qy 2574 GTTAAAAAGATAAAATTTGAGCTGCACCAGGCTGAGGAGCTTAAATGCTCTTGGAAAAAGGAA 2633

Db 4173 GCTGAGGCATCACTATGAGCAACAGCAGCAGGAAGACCTAGCCCGGAATCTAGGCAAGGGCAA 4232

Qy 2634 GAGAACTCGCAAGCAGTTGGTATCCATTGAAGAAAGATGATCTTGTGCTTGTGGAAAGATGT 2693

Db 4233 GCGGTTTCGCAAGCAAGTTAACTACATATGATGCTGCTCAGGAAGACCAAGACAACCATGTC 4292

Qy 2694 GAGCTCTGATGGAGATGAAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGACA 2753

Db 4293 AGAGTACTCGTGGGTTTCAGAGGAGGAGATGAAGACTTCGATGAACGCTCCTGAAGGGCG 4352

Qy 2754 AGGAGTTCAGACGGGTCGACGGCCGTACAGAAAGAAAGGTCGCGATAAATTTGGAACCAAC 2813

Db 4353 TAGACAGTCAAAGAGGCAGCTCCGAA-----TGAGAAAGATAAGCCACTGCC 4400

Qy 2814 TCCGTTGATGGAAGGTGAGGGGAGATCTTTTCAGAGTACTGGGTTTCAACAGAGTCAAAAG 2873

Db 4401 TCCACTGCTGGCCCGAGTCCGGGGCAACATTTGAGGTGCTGGGCTTCAACACCCGTCAGCG 4460

Qy 2874 GGCCATTTTGTACAGACTTTTGTAGAGGTATGG-----AGCTGGCAATTTTGATTG 2924

Db 4461 GAAGGCTTTCTCAATGCTGTGATCGCTGGGGATGCCACCACAGGATGCCTTCACCAC 4520

Qy 2925 GAAGGAGTTTGTCTCCTCGCTTAAAGCAGAAAGACCTTTGAAGAAATAAATGAATATGGAAT 2984

Db 4521 ACAGTGGCTGTGCGGGACCTGAGGGGCAAGACTGAGAAGGAGTTTAAGGCCTATGTGTC 4580

Qy 2985 ACTCTTCTTGAAGCACATTTGCTGAAGAAATAGACGAGAAATTTCCAAACCTTTTCAGATGG 3044

Db 4581 TTTGTTTCATGCGCCATCTGTGTGAGCCCTGGGGCAGACGGCTCTGAAAACCTTTGCCGATGG 4640

Qy 3045 TGTGCCAAAGGAAGGACTTTAGAATAGAAAGATGTTCTAGTCAAGAAATGCTCTTCTGATACT 3104

Db 4641 GGTCCTCGGGAGGAGCTGAGTCGCCAGCAGGTTTGAACCCGATTTGGAGTCATGTCTCT 4700

Qy 3105 AGTTTCAGGAGAAGGTGAA 3122

Db 4701 CGTCAAAAAGAGGTGCA 4718

RESULT 13

AAD54629

ID AAD54629 standard; DNA; 6331 BP.

XX

AC AAD54629;

XX

DT 26-JUN-2003 (first entry)

XX

DE Human chromodomain helicase DNA binding protein (CHD) encoding DNA #4.

XX

KW Human; p53 pathway; therapeutic; angiogenic disorder; apoptotic disorder; chromodomain helicase DNA binding protein; CHD; cancer; gene therapy; cell proliferative disorder; chromatin organisation modifier domain; cytostatic; ds.

XX

OS Homo sapiens.

XX

PN WO200298899-A2.

XX

PD 12-DEC-2002.

XX

PF 03-JUN-2002; 2002WO-US017466.

XX



PR 05-JUN-2001; 2001US-0296076P.  
PR 10-OCT-2001; 2001US-0328605P.  
PR 22-OCT-2001; 2001US-0338733P.  
PR 15-FEB-2002; 2002US-0357253P.  
PR 15-FEB-2002; 2002US-0357600P.  
XX  
PA (EXEL-) EXELIXIS INC.  
XX  
PI Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;  
PI Lioubin MN;  
XX  
XX WPI; 2003-156840/15.  
DR  
XX  
PT Identifying a candidate p53 pathway-modulating agent as therapeutic  
PT targets for disorders related to defective p53 function e.g. cancer by  
PT contacting an assay system having purified CHD polypeptide or nucleic  
PT acid, with a test agent.  
XX  
PS Disclosure; Page 51-54; 278pp; English.  
XX  
CC The present invention relates to a method for identifying candidate p53  
CC pathway modulating agents. The method involves contacting an assay system  
CC comprising purified chromatin organisation modifier (chromo) domain  
CC helicase DNA binding proteins (CHD), nucleic acids, their functionally  
CC active fragments or derivatives, with a test agent under conditions  
CC where, but for the presence of the test agent, the system provides a  
CC reference activity. The methods are useful for identifying modulators of  
CC the p53 pathway as therapeutic targets for disorders associated with  
CC defective p53 function, such as angiogenic disorders, apoptotic disorders  
CC or cell proliferative disorders, e.g. cancer. The modulators are useful  
CC as research reagents, diagnostics and therapeutics. The invention is also  
CC useful in gene therapy. The present sequence is human CHD DNA  
XX  
SQ Sequence 6331 BP; 1720 A; 1552 C; 1819 G; 1240 T; 0 U; 0 Other;  
  
Query Match 10.1%; Score 420.8; DB 7; Length 6331;  
Best Local Similarity 52.3%; Pred. No. 2e-101;  
Matches 1233; Conservative 0; Mismatches 1032; Indels 93; Gaps 10;  
  
QY 810 AGGCTTGTTACATCCATACCAGCTTGAGGGACTTAATTTTGGGGTTCTCGTGGTCAA 869  
DB 2409 AGGCACCTGCACATGTATCAGTTGGAAGGGCTGAAGTGGTACGCTTCTCTGGGCCA 2468  
  
QY 870 ACAGACGCATGTAATCCTTGCTGATGAAATGGGATAGGCAAGACAAATTCAAAGCATTGC 929  
DB 2469 GGGCACTGACACCAATCTAGCTGATGAGATGGGGCTAGGCAAGCACCATAACAACCATCGT 2528  
  
QY 930 CCTTTTAGCTTCACTTTTGGAGGAACTCA-----TTCCGCATTTGGTAATTGCTCC 983  
DB 2529 CTTCCTCTACTCACTCTACAAGGAGGGCCACACAAAAGGTCCTTCTCTGGTGGTCCCC 2588  
  
QY 984 TCTATCGACTCTCGGTAACCTGGGAGAGAGAGTTTGGCCACATGGGCCCCACAGATGAACGT 1043  
DB 2589 ACTCTCTACCATCAATTAATGGGAGCGGAGTTCAGATGTGGGCACCCCAATCTATGT 2648  
  
QY 1044 GGTATGTATTTTGGCACTGCGCAAGCTCGAGCAGTATATCAGAGAACATGAGTTTACTT 1103  
DB 2649 GGTGACATACA CGGGTGACAAGGACAGCCGGGCCATCATTCGTGAGAAATGAATTCCTT 2708  
  
QY 1104 ATCGAAAGATCAAAAAAGATCAAGAAAAAGAAATCTGGACAAATAAGTAGCGAAAGCAA 1163  
DB 2709 TGAGGACAAATGCCATCAAGGGGGCAAGAAAGCTTTTAAAGATGAAAAG----- 2756  
  
QY 1164 GCAAAAAAGAAATCAAGTTTGATGTCTCCTCACATCGTATGAGATGATCAACCTAGATTTC 1223  
DB 2757 GGAGGCACAGGTGAAGTTCCATGTCTCCTTGACATCGTATGAGCTGATCACCATTGATCA 2816  
  
QY 1224 ACAGTTCTAAACCAATTAAGTGGGAGTGCATGATGTTGATGAAGGTCAATCGACTGAA 1283  
DB 2817 GGCAGCACTTGGTTCCATCGCTGGGCCCTGTCTTGTGGTAGATGAGGCCCATCGACTCAA 2876  
  
QY 1284 AAATAAGGATTCAAAGCTGTCTCTTCAATTGACACAGTATTCAAGTAACCAACCGTATTCT 1343  
DB 1343

DB 2877 GAACAACCAAGTCTTTCAGGGTTCTCAATGGTTACAAGATAGATCATAAAGTTGCT 2936  
QY 1344 TCTGACAGGAACACCACTTTCAGAACAACTTGGATGAACCTTTTCATGCTCATGCTTTTCT 1403  
DB 2937 GCTGACAGGAACCCCATTTGCAGATAAATCTGGAGGAGCTCTTCCATCTCTGAACCTTCT 2996  
QY 1404 TGATGCGGGGAAGTTTGGAAAGTTTGGAGGAGTTCCAGGAGGAGTTCAAAAGATATTAAATCA 1463  
DB 2997 CACCCACAGAGAGATTTAACAACCTTGGAGGGCTTCTTGGAGGAGTTTGTGACATATCCAA 3056  
QY 1464 AGAGGACAGATCTCAAGGTTTGCACAAAATGTTGGCTCCACATTTTCTCAGAAAGGTAAA 1523  
DB 3057 AGAGGACAGATCAAGAAACTGTCATGATTTGCTGGGGCCACACATGCTCGGAGACTCAA 3116  
QY 1524 AAAAGAGTAATGAAAGACATGCCCCCAAAAAGGAGCTCATTTTGGCGTGTGATCTGAG 1583  
DB 3117 GGCAGATGTCTTTAAGAACATGCCAGCCAAAGACAGAGCTCATCGTTCGGGTGGAGCTAAG 3176  
QY 1584 CAGTCTGCAGAAAAGAAATATTACAAAGCTATTTTACCCGTAATATATCAAGTATTGA--C 1640  
DB 3177 CCCCATGCAGAAAGAAATACTACAAATACATCTGACTCGAAATTTTGGAGCCCTTGAATTC 3236  
QY 1641 AAAAAAGGAGGTGCTCAAAATTTCCCTTAATAACATTTATGATGAATTAAGAAAAGTATG 1700  
DB 3237 ACGAGGTGTTGGAAACCAGGTGTCGCTGCTTAATATCATGATGATCTTAAGAAGTGTCTG 3296  
QY 1701 CTGCCATCCTTATATGCTAGAGGGGTGTTG-----AGCCAGTTATTACGACGCAAA 1751  
DB 3297 CAACCATCCATACCTTTTCCCGTGGTGTCTATGGAGTCCCCCAAACTCCCCAGTGGGC 3356  
QY 1752 TGAAGCTTTCAAAACAACCTTTTGGAGTCTTGTGGAAAGCTGCAACTTCTAGATAAAATGAT 1811  
DB 3357 TTATGAGGTGGGGCACCTTATTAAGTCGCTCTGGGAAGCTCATGCTGCTCCAGAAGATGCT 3416  
QY 1812 GGTCAAACTGAAAGAGCAAGGACACAGAGTCCCTAATATACACAGTTTCAGCATATGCT 1871  
DB 3417 GCGAAAGCTGAAGGAGCAAGGACACCGAGTGTCTCATCTTCTCGAGATGACCAAAATGTT 3476  
QY 1872 GGACTTACTTGAAGACTACTGTACCATAAAGAAATGGCAGTACGAGCGAAATTTGATGGAAA 1931  
DB 3477 AGACTTGTCTGAGGACTTCTTAGACTATGAAGGCTACAAGTATGAGCGCATCGATGTGTG 3536  
QY 1932 GGTGCGGAGCTGAGCGGCAAAATACGCATAGATCGGTTCAATGCAAAAAAATTTCTAACAA 1991  
DB 3537 TATCAGGGTGCCTTGAGGCAAGGAGGCGCATCGATCGGTTTAAATGCTCTGGGGCCCAACA 3596  
QY 1992 GTTTTGTGTTTGTCTCTCCACAAGAGCTGGTGGCTTAGGAATAAATCTTCAACGGCTGA 2051  
DB 3597 ATCTGTCTTCTCTGTCCACCCGAGCTGGGGCCCTGGGCATCAATCTGGCCACTGCTGA 3656  
QY 2052 TACAGTAATCATTTATGACAGTGAATCCTCATGCTGATCTTCAAGCAATGGCTAG 2111  
DB 3657 CACTGTCAATCATCTTTGATTCTGACTGGAACCCCAATAATGACATCCAGGCCCTTTAGCCG 3716  
QY 2112 AGCTCATCGACTTGGCCAAAACAAATAAGGTGATGATTTATAGGCTCATAAACCGAGGCAC 2171  
DB 3717 GGCTCATCGGATTTGGCCAGGCCAACAAAGTGAATTTACCGGTTTGTGACTCGCGCGTC 3776  
QY 2172 CATTGAAGAAAGGATGATGCAATTTGACTTAAAAAGAAAAATGGTTCTAGAGCATCTTGTGT 2231  
DB 3777 AGTGAAGAGCGGAATCACACAAGTGGCCAAAGAGAAAAGATGATGCTGACACACCTGTTGT 3836  
QY 2232 TGG-----GAAACTCAAAACACAAAAACATTAATCAGGAAGAGTTAGATGACATCAT 2282  
DB 3837 GCGGCCTGGGCTGGGCTCCAAGGCAAGGCTCCATGTCCAAGCAGGAGCTTGACGACATTTCT 3896  
QY 2283 CAGGTATGGATCAAAAGGAGCTTTTGTGCTAGTGAAGATGATGAAGCAGGAAAGTCTG---- 2338  
DB 3897 CAAATTTGGCACTGAAGAGCTATTCAAGGATGAAAACGAGGGGGAGAACAAAGGAGGAGGA 3956  
QY 2339 -----GAAAAATTCAATTATGATGATGCGGCTATAGACAAATTTGCTTGTGATCTCGT 2393  
DB 3957 CAGCAGTGTGATTCAATTATGACAATGAGGCCATCGCTCGGCTGTTGGACCGGAACCCAGGA 4016

QY 2394 GGAGGCAGAGGAAGTCTCAGTGGATGATGAAGAGGAGAAATGGATTCTTAAAGGCTTTCAA 2453  
| | | | |  
Db 4017 TGCAACTGAGGA-----CACTGACGTGCAGAACATGAATGAGTATCTCAGCTCCTTTCAA 4070  
  
QY 2454 GGTGGCTAATTTGAATATATAGATGAATAATGAGGCAGCAGCAATTAGAGGCACAGAGAGT 2513  
| | | | |  
Db 4071 GGTGGCACAGTACGTCGTGCGGGAAGAGACAAGATTGAGGAAATTGAGCGAGAGATCAT 4130  
  
QY 2514 CGCTGCTGAAAGCAAATCTTCAGCAGGCAATCTGTATAGAGCAAGTTATTTGGGAAGAGTT 2573  
| | | | |  
Db 4131 CAAGCAGGAGGAGAAATGTGACCCCTGAC-----TACTGGGAGAAGCT 4172  
  
QY 2574 GTTAAAGATAAATTTGAGCTGCACCAGGCTCAGGAGCTTAATGCTCTTGGAAAAAGGAA 2633  
| | | | |  
Db 4173 GCTGAGGCATCACTATGAGCAACAGCAGGAGACCTAGCCCGAATCTAGGCAAGGGCAA 4232  
  
QY 2634 GAGAAAGTCGCAAGCAGTTGGTATCCATTGAAGAAGATGATCTTGGCTGGTTTGAAGATGT 2693  
| | | | |  
Db 4233 GCGGTTTCGCAAGCAAGTTAACTACAATGATGCTCTCAGGAAGACCAAGACCAACCGTC 4292  
  
QY 2694 GAGCTCTGATGAGATGAAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGACA 2753  
| | | | |  
Db 4293 AGAGTACTCGTGGGTTCAGAGGAGGAGATGAAGACTTTCGATGAACGTCCTGAAGGGCG 4352  
  
QY 2754 AGGAGTTCAGACGGGTCGACGGCCGTACAGAAGAAAGGGTCGCGAATAATTGGAACCAAC 2813  
| | | | |  
Db 4353 TAGACAGTCAAAAGAGGCAGCTCCGGAA-----TGAGAAAGATAAGCCACTGCC 4400  
  
QY 2814 TCCGTTGATGGAAGGTGAGGGGAGATCTTTTCAGAGTACTGGGTTTCAACCAGAGTCAAAG 2873  
| | | | |  
Db 4401 TCCACTGCTGCGCCGAGTCGGGGCAACATGAGGTGCTGGGCTTCAACACCCGTCAGCG 4460  
  
QY 2874 GGCCATTTTGTACAGACTTTGATGAGGATGG-----AGCTGGCAATTTTGATTG 2924  
| | | | |  
Db 4461 GAAGGCTTTTCTCAATGCTGTGATGCGTGGGGATGCCACCACAGGATGCCCTTCACCAC 4520  
  
QY 2925 GAAGGAGTTTGTCTCCTCGCTTAAAGCAGAGACCTTTGAAGAAATAAATGAATATGGAAT 2984  
| | | | |  
Db 4521 ACAGTGGCTGTGTCGGGACCTGAGGGGCAAGACTGAGAAGGAGTTTAAAGGCCATATGTGC 4580  
  
QY 2985 ACTCTTCTTGAGCACATTTGCTGAAGAAATAGACGAGAATTTCTCCAACCTTTTCAGATGG 3044  
| | | | |  
Db 4581 TTTGTTTCATGCGCCATCTGTGTGAGCCTGGGSCAGACGGCTCTGAACCTTTGCCGATGG 4640  
  
QY 3045 TGTGCCCCAAGGAAGGACTTAGAATAGAAAGATGTTCTAGTCAGAATTGCTTCTTGATACT 3104  
| | | | |  
Db 4641 GGTCCCTCGGGAGGAGTGAAGTCGCCAGCAGGTGTTGACCCGATTCAGGATCATGTCTCT 4700  
  
QY 3105 AGTTCAGGAGAGGTGAA 3122  
| | | | |  
Db 4701 CGTCAAAAAGAAAGGTGCA 4718

RESULT 14  
ACD13374  
ID ACD13374 standard; cdna; 6331 BP.  
XX  
AC ACD13374;  
XX  
DT 13-AUG-2003 (first entry)  
XX  
DE Human DNA encoding a p53 modifier, SEQ ID 44.  
XX  
KW Human; ss; gene; p53 modifier; cytostatic; cancer; cytostatic;  
KW antiangiogenic; antiapoptotic; p53 pathway; breast cancer; colon cancer;  
KW kidney cancer; lung cancer; ovarian cancer; angiogenesis; cell cycle;  
KW apoptotic disorder; cell proliferation disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200299122-A1.  
XX

PD 12-DEC-2002.  
XX  
PF 03-JUN-2002; 2002WO-US017382.  
XX  
PR 05-JUN-2001; 2001US-0296076P.  
PR 10-OCT-2001; 2001US-0328605P.  
PR 15-FEB-2002; 2002US-0357253P.  
XX  
PA (EXEL-) EXELIXIS INC.  
XX  
PI Friedman L, Flowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;  
XX  
DR WPI; 2003-156859/15.  
DR P-PSDB; ABO07199.  
XX  
PT Identifying modulators of the p53 pathway for use in treating apoptotic  
or cell proliferation disorders, comprises screening for agents that  
modulate activity of a human ortholog of genes that modify the p53  
pathway in Drosophila.  
XX  
PS Example 2; Page 269-271; 678pp; English.  
XX  
CC The invention relates to identifying (M1) a candidate p53 pathway  
modulating agent, by contacting an assay system comprising a purified HM  
polypeptide (human orthologue of genes that modify the p53 pathway in  
Drosophila) or nucleic acid with a test agent under conditions, where but  
for the presence of the test agent, the system provides a reference  
activity, and detecting a test agent-biased activity of the assay system.  
CC Also included are modulating (M2) a p53 pathway of a cell (comprising  
contacting a cell defective in p53 function with a candidate modulator  
that specifically binds to a HM polypeptide comprising an HM amino acid  
sequence, where p53 function is restored), modulating (M3) a p53 pathway  
in a mammalian cell (comprising contacting the cell with an agent that  
specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4)  
a disease in a patient (comprising: (a) obtaining a biological sample  
from the patient; (b) contacting the sample with a probe for HM  
expression; (c) comparing the results with a control; and (d) determining  
whether the comparison indicates a likelihood disease). (M1) is useful  
for identifying modulators of the p53 pathway. A probe for HM expression  
is useful for diagnosing breast, colon, kidney, lung and ovarian cancer,  
in a patient, where the cancer has greater than 25 % expression level.  
CC Modulators identified by (M1) are useful in a variety of diagnostic and  
therapeutic applications, where disease or disorder prognosis is related  
to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell  
proliferation disorders (e.g. cancer). Another two new methods (M2 and  
M3) are useful for modulating the p53 pathway of a cell, thus restoring  
the p53 function of the cell, so that the cell undergoes normal  
proliferation or progression through the cell cycle. (M2) and (M3) are  
also useful for treating defects in the p53 pathway such as angiogenic,  
apoptotic or cell proliferation disorders. The present sequence is an HM  
nucleic acid encoding a p53 pathway modifying protein  
XX  
SQ Sequence 6331 BP; 1720 A; 1552 C; 1819 G; 1240 T; 0 U; 0 Other;

Query Match 10.1%; Score 420.8; DB 7; Length 6331;  
Best Local Similarity 52.3%; Pred. No. 2e-101;  
Matches 1233; Conservative 0; Mismatches 1032; Indels 93; Gaps 10;  
  
QY 810 AGGCTTGTACATCCATACCAGCTTGAGGGACTTAATTTTTCGGTTCTCGTGGTCAA 869  
| | | | |  
Db 2409 AGGCACCCTGCACATGTATCATGTTGGAAGGGCTGAACTGGTACGTTCTCTGGGCCCA 2468  
  
QY 870 ACAGACGCATGTAATCCTTGTGATGAATAATGGGACTAGGCAAGACAATTCAAAGCATTCG 929  
| | | | |  
Db 2469 GGGCACTGACACCATTTCTAGCTGATGAGATGGGGCTAGGCAAGACCATACAAAACCATCGT 2528  
  
QY 930 CCTTTTAGCTTCACTTTTGTAGGAGAACCTCA-----TTCCGCAATTTGGTAATTGCTCC 983  
| | | | |  
Db 2529 CTTCTCTACTCACTCTACAAGGAGGGCCACACAAAAGGTCCTTCTCTGGTGGTGGCCCC 2588  
  
QY 984 TCTATCGACTCTGCGTAACCTGGGAGAGAGAGTGTTCACACATGSGGCCCCACAGATGAACGT 1043  
| | | | |  
Db 2589 ACTCTCTACCATCATTAACCTGGGAGCGGGAGTTCACAGATGTGGSCACCCAAATTCATGT 2648





RESULT 15  
ABX63370  
ID ABX63370 standard; cDNA; 7805 BP.  
XX  
AC ABX63370;  
XX  
DT 25-FEB-2003 (first entry)  
XX  
DE Human cDNA #370 differentially expressed in activated vascular tissue.  
XX  
KW Human; gene; ss; vascular tissue; cytostatic; atherosclerosis; cardiant;  
KW hypotensive; antidiabetic; gynaecological; vasotropic; cerebroprotective;  
KW gene therapy; vascular disease; cancer; coronary; artery disease;  
KW hypertension; diabetes; pre-eclampsia; restenosis;  
KW ischaemia-reperfusion injury; stroke.  
XX  
OS Homo sapiens.  
XX  
PN US2002137081-A1.  
XX  
PD 26-SEP-2002.  
XX  
PF 08-JAN-2002; 2002US-00044090.  
XX  
PR 28-JUL-2000; 2000US-0222469P.  
PR 08-JAN-2001; 2001US-0260483P.  
XX  
PA (BAND/) BANDMAN O.  
XX  
PI Bandman O;  
XX  
DR WPI; 2003-110597/10.  
XX  
PT Combination for diagnosing, staging, treating, or monitoring the  
PT progression of treatment of a vascular disease, e.g. atherosclerosis,  
PT comprises several cDNAs that are differentially expressed in activated  
PT vascular tissue.  
XX  
PS Claim 1; Page; 18pp; English.  
XX  
CC This invention relates to a combination comprising several cDNAs that are  
CC differentially expressed in activated vascular tissue. The invention also  
CC discloses a high throughput method for detecting differentially expressed  
CC cDNAs in a sample. The cDNAs of the invention may have  
CC antiarteriosclerotic; cytostatic; cardiant; hypotensive; antidiabetic;  
CC gynaecological; vasotropic and cerebroprotective activities and may be  
CC used in gene therapy. The cDNAs of the invention may be used in a high-  
CC throughput methods for detecting differential expression of one or more  
CC cDNAs in a sample, or screening several molecules or compounds to  
CC identify a molecule or compound that specifically binds a cDNA of the  
CC invention. A protein encoded by the cDNA may be used to screen several  
CC molecules or compounds to identify a ligand that specifically binds to  
CC the protein, or to produce or purify an antibody to the protein that can  
CC be used to detect a protein in a sample or purify a natural or  
CC recombinant protein from a sample. The nucleotides may be useful for  
CC diagnosing, staging, treating, or monitoring the progression of treatment  
CC of a vascular disease, e.g. atherosclerosis, cancer, coronary artery  
CC disease, hypertension, diabetes, pre-eclampsia, ischaemia- reperfusion  
CC injury, restenosis, or stroke. The cDNAs can also be used for large-scale  
CC genetic or gene expression analysis of several new nucleic acid  
CC molecules. Antibodies to the proteins encoded by the cDNAs are useful for  
CC diagnosing pre-pathologic disorders, and chronic or acute diseases  
CC associated with abnormalities in the expression, amount or distribution  
CC of the protein. The present sequence represents a cDNA of the invention  
CC that is differentially expressed in activated vascular tissue. Note: The  
CC sequence data for this patent did not form part of the specification, but  
CC was obtained in electronic format directly from USPTO at  
CC <http://segdata.uspto.gov/sequence.html?DocID=20020137081>  
XX  
SQ Sequence 7805 BP; 2003 A; 1945 C; 2323 G; 1534 T; 0 U; 0 Other;

Query Match 9.8%; Score 409.6; DB 7; Length 7805;  
Best Local Similarity 52.5%; Pred. No. 2.2e-98;

Matches 1240; Conservative		0;	Mismatches 1024;	Indels	96;	Gaps	12;
QY	810	AGGCTTGTTACATCCATACCAGCTTGAGGACTTAATTTTTTTCGGTCTCTCGTGGTCAAA	869				
Db	2692	AGGCACCTGACATGTATCAGTTGGAAGGGCTGAACTGGCTACGCTTCTCTGGGCCCA	2751				
QY	870	ACAGACGCTGTAAATCCTTGCTGATGAAATGGGACTAGGCAAGACAAATTCAAAGCATTC	929				
Db	2752	GGGCACTGACACCATTTCTAGCTGATGATGGGCTAGGCAAGACCAATACAAACCATCGT	2811				
QY	930	CTTTTGTAGCTTCACTTTTGTAGGAGAACCTCA-----TTCCGCAATTTGGTAAATGCTCC	983				
Db	2812	CTTCTCTACTCACTCTACAAGGAGGGCCACACAAAGGTCCTTCTCTGTGAGTGCCCC	2871				
QY	984	TCTATCGACTCTGCGTAACTGGGAGAGAGAGTTTGCCACATGGGCCCCACAGATGAACGT	1043				
Db	2872	ACTCTTACCATCATTAATCTGGAGCGGAGTTCCAGATGTGGGCACCCAAATTTCTATGT	2931				
QY	1044	GGTTATGTATTTGGCACTGCGCAAGCTCGACAGTTATCAGAGAACATGAGTTTTACTTT	1103				
Db	2932	GGTGACATACACGGGTGACAAGGACAGCCGGGCCCATCATTCGTGAGATGAATTTCTCCTT	2991				
QY	1104	ATCGAAAGAT-CAAAAAAGATCAAGAAAAAGAAATCTGGACAAATAAGTAGCGAAAGCA	1162				
Db	2992	TGAGGACAAATGCCATCAAAGGGGGCAAGAAAGCTTTTAAAGATGAAGGGAGGCACAGG-	3050				
QY	1163	AGCAAAAAAGAAATCAAGTTTGATGTCCTCCTCACATCGTATGAGATGATCAACCTTAGATT	1222				
Db	3051	-----TGAAGTTCCATGTTCTCTCGACATCGTATGAGCTGATCACCATTGATC	3098				
QY	1223	CAGCAGTTCTAAAAACCAATTAAAGTGGGAGTGCATGATTGTTGATGAAGGTTCATCGACTGA	1282				
Db	3099	AGGCAGCACTTGGTTCCATCCGCTGGGCCCTGTCTTGTGGTAGATGAGGCCCATCGACTCA	3158				
QY	1283	AAAATAAGGATTCAAAGCTGTTCTCTTCATTGACACAGTATTCAAGTAACCAACCGTATTC	1342				
Db	3159	AGAACAAACAGTCCAAGTTTTTTCAGGGTTCTCAATGGTTACAAGATAGATCAAGTTGC	3218				
QY	1343	TTCTGACAGGAACACCACTTCAGAACAACTTGGATGAACCTTTTTCATGCTCATGCAATTTTC	1402				
Db	3219	TGCTGACAGGAACCCCATTCGAGAATAATCTGGAGGAGCTCTTCCATCTCCTGAACTTCC	3278				
QY	1403	TTGATCGGGGAAGTTTGGAAAGTTTGGAGGAGTTCAGGAGGAGTTCAAAAGATATTAATC	1462				
Db	3279	TCACCCCAAGAGAGATTAAACAACCTTGGAGGCTTCTTGGAGGAGTTTGTGACATATCCA	3338				
QY	1463	AAGAGGACGAGATCTCAAGTTTGCACAAAATGTTGGCTCCACATTTGCTCAGAAGGGTAA	1522				
Db	3339	AAGAGGACCAAGATCAAGAAACTGCAATGATTTGCTGGGCCACACATGCTGCGGAGACTCA	3398				
QY	1523	AAAAAGACGTAATGAAAGACATGCCCCCCCCAAAAAGGAGCTCATTTTTCGCTGTGATCTGA	1582				
Db	3399	AGGCAGATGTTTAAAGAACATGCCAGGCCAAGACAGAGCTCATCGTTTCGGGTGGAGCTAA	3458				
QY	1583	GCAGTCTGCAGAAAGAAATATTACAAAGCTATTTTTTACCCGTAATTATCAAGTATTGA--	1639				
Db	3459	GCCCCATGCAGAAAGAAATACTACAAATACATCCTGACTCGAAATTTTGGGCTTGAATT	3518				
QY	1640	CAAAAAAGGAGGTGCTCAAAATTTCCCTTAAATAAACATTAATGATGGAATTAAGAAAAATAT	1699				
Db	3519	CACGAGGTGCTGGGAACCAAGGTGTCGCTGCTTAATATCATGATGGATCTTAAGAAGTGCT	3578				
QY	1700	GCTGCCATCCTTATATGCTAGAGGTGTTG-----AGCCAGTTATTACACGACGCAA	1750				
Db	3579	GCAACCATCCATACCTTTTTTCCCGTGGCTGCTATGGAGTCCCCCAAACTCCCCAGTGGG	3638				
QY	1751	ATGAAGCTTTCAAAACAACCTTTTGGAGTCTTGTGGAAAGCTGCAACTTTCTAGATAAAATGA	1810				
Db	3639	CTTATGAGGGTGGGGCACTTATTAAAGTCGTCTGGGAAGCTCATGCTGCTCCAGAAGATGC	3698				
QY	1811	TGGTCAAACTGAAAGAGCAAGGACACAGAGTCTCTAATAATATACACACAGTTTCAGCATATGC	1870				
Db	3699	TGCGAAAGCTTGAAGGAGCAAGGACACCGAGTGTCTCATCTTCTCGAGATGACCAAAATGT	3758				



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 15:45:46 ; Search time 266 Seconds  
(without alignments)  
8714.398 Million cell updates/sec

Title: US-10-049-137-1  
Perfect score: 4177  
Sequence: 1 atgagtagtttggtggagag.....caactggtaaatcaagattc 4177

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/pCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	421.4	10.1	6328	4 US-08-913-832A-1	Sequence 1, Appli
2	421.4	10.1	6328	4 US-09-249-181A-1	Sequence 1, Appli
3	421.4	10.1	6328	4 US-09-158-707-1	Sequence 1, Appli
4	421.4	10.1	6475	4 US-09-620-312D-325	Sequence 325, App
5	285.4	6.8	5837	4 US-09-976-594-75	Sequence 75, Appl
6	213.6	5.1	2512	4 US-09-702-705-1802	Sequence 1802, Ap
7	213.6	5.1	2512	4 US-09-736-457-1802	Sequence 1802, Ap
8	213.6	5.1	2512	4 US-09-671-325-1802	Sequence 1802, Ap
9	112.2	2.7	8916	4 US-09-579-181-11	Sequence 11, Appl
10	112.2	2.7	9354	4 US-09-579-181-10	Sequence 10, Appl
11	103.8	2.5	1868	4 US-09-023-655-429	Sequence 429, App
12	103.2	2.5	575	3 US-09-328-111-553	Sequence 553, App
13	100.2	2.4	5386	4 US-09-535-008-60	Sequence 60, Appl
14	100.2	2.4	5468	4 US-09-535-008-66	Sequence 66, Appl
15	100.2	2.4	5471	4 US-09-535-008-1	Sequence 1, Appli
16	100.2	2.4	5471	4 US-09-535-008-62	Sequence 62, Appl
17	100.2	2.4	5477	4 US-09-535-008-74	Sequence 74, Appl
18	100.2	2.4	5480	4 US-09-535-008-70	Sequence 70, Appl
19	100.2	2.4	5564	4 US-09-535-008-68	Sequence 68, Appl
20	100.2	2.4	5567	4 US-09-535-008-64	Sequence 64, Appl
21	100.2	2.4	5573	4 US-09-535-008-76	Sequence 76, Appl
22	100.2	2.4	5576	4 US-09-535-008-72	Sequence 72, Appl
23	90.6	2.2	1220	4 US-09-205-258-140	Sequence 140, App
24	89.2	2.1	3436	3 US-09-276-531-60	Sequence 60, Appl
25	88.2	2.1	345	4 US-09-016-434-15	Sequence 15, Appl
26	79.8	1.9	3593	4 US-09-404-627-3	Sequence 3, Appli
27	79.8	1.9	4205	4 US-09-404-627-1	Sequence 1, Appli

C 28	79.2	1.9	535	4 US-09-389-681-291	Sequence 291, App
C 29	79.2	1.9	535	4 US-09-620-405B-291	Sequence 291, App
C 30	79.2	1.9	535	4 US-09-339-338-291	Sequence 291, App
C 31	79.2	1.9	535	4 US-09-433-826B-291	Sequence 291, App
C 32	79.2	1.9	535	4 US-09-604-287A-291	Sequence 291, App
C 33	79.2	1.9	535	4 US-09-834-759-291	Sequence 291, App
C 34	77.2	1.8	737	3 US-08-998-416-851	Sequence 851, App
C 35	73.2	1.8	11864	4 US-08-961-527-61	Sequence 61, Appl
C 36	70.8	1.7	580073	4 US-08-545-528D-1	Sequence 1, Appli
C 37	70	1.7	7218	1 US-08-232-463-14	Sequence 14, Appl
C 38	69.6	1.7	1230025	4 US-09-198-452A-1	Sequence 1, Appli
C 39	65.8	1.6	3213	4 US-09-107-532A-62	Sequence 62, Appl
40	65.8	1.6	1230025	4 US-09-198-452A-1	Sequence 1, Appli
41	65.2	1.6	656	3 US-08-998-416-974	Sequence 974, App
42	60	1.4	5000	3 US-09-104-070-1	Sequence 1, Appli
43	57	1.4	239	4 US-09-016-434-267	Sequence 267, App
44	54.2	1.3	253	4 US-09-016-434-465	Sequence 465, App
45	53.6	1.3	293	4 US-09-016-434-897	Sequence 897, App

ALIGNMENTS

RESULT 1  
US-08-913-832A-1  
; Sequence 1, Application US/08913832A  
; Patent No. 6329517  
; GENERAL INFORMATION:  
; APPLICANT: Seelig, Hans Peter  
; APPLICANT: Renz, Manfred  
; TITLE OF INVENTION: DERMATOMYOSITIS-SPECIFIC AUTO-ANTIGEN  
; FILE REFERENCE: 8484-0030-999  
; CURRENT APPLICATION NUMBER: US/08/913,832A  
; CURRENT FILING DATE: 1998-01-12  
; PRIOR APPLICATION NUMBER: PCT/DE96/00444  
; PRIOR FILING DATE: 1996-03-08  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 6328  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(5736)  
US-08-913-832A-1

Query Match	10.1%;	Score 421.4;	DB 4;	Length 6328;
Best Local Similarity	51.9%;	Pred. No. 1.2e-111;		
Matches 1236;	Conservative	0;	Mismatches 1076;	Indels 69; Gaps 10;
QY	818	TACATCCATACCAGCTTGAGGGAGCTTAATTTTTCGGGTTCTCGTGGTCAAACAGACGC	877	
Db	2177	TGCACCCCTATCAAATGGAGGCGCTGAATGGTTGCGCTTCTCTGGGCTCAGGGCACTG	2236	
QY	878	ATGTAATCCTTGCTGATGAAATGGGACTAGGCAAGACAATTCAAAGCATTCGCCCTTTTAG	937	
Db	2237	ACACCATCTGGCTGATGAGATGGGCTTGGGAAACTGTACAGACAGCAGTCTTCCTGT	2296	
QY	938	CTTCACCTTTTGAGGAG-----AACCTCATTCGCGCATTTGGTAATTGCTCCTCTATCGA	991	
Db	2297	ATTCCCTTTACAAGGAGGGTCAATCCAAAGGCCCTTCCTAGTAGCGGCCCTCTTTCTA	2356	
QY	992	CTCTGCGTAACCTGGGAGAGAGAGTTTGCCACATGGGCCCCCAGACATGAACGTGGTTATGT	1051	
Db	2357	CCATCATCAACTGGGAGCGGGAGTTTGAAATGTGGGCTCCAGACATGTATGTCGTAACCT	2416	
QY	1052	ATTTTGGCACTGCGCAAGCTCGAGCAGTTATCAGAGAACTAGAGTTTACTTATCGAAAG	1111	
Db	2417	ATGTGGGTGACAAAGACAGCCGTGCCCATCATCCGAGAGAAATGAGTTCTCCTTTGAAGACA	2476	
QY	1112	ATCAAAAAAAGATCAAGAAAAAGAAATCTGGACAAATAAGTAGCGAAAGCAAAAAA	1171	



Db 2477 ATGCCATTCTGGTGGCAAGGCGCTC-----CCGCATGAAGAAAGAGGCAT 2524

QY 1172 GAATCAAGTTTGTATGTCCTCTCACATCGTATGAGATGATCAACCTAGATTTCAGCAGTTC 1231

Db 2525 CTGTGAAATTCATGCTGCTGACATCCTATGAATTGATCACCATTGACATGGCTATTT 2584

QY 1232 TAAACCAATTAAGTGGAGTGCATGATTGTTGATGAAGGTTCATCGACTGAAAAATAAGG 1291

Db 2585 TGGGCTCTATTGATTGGGCTGCCTCATCGTGGATGAAGCCCATCGGCTGAAGAACATC 2644

QY 1292 ATTCAAAGCTGTTCTTCATTGACACACAGTATTCAAAGTAACCAACCGTATCTTCTGACAG 1351

Db 2645 AGTCTAAGTTCTTCCGGGTATTGAATGGTTACTCACTCCAGCACAAAGCTGTTGCTGACTG 2704

QY 1352 GAAACACCACCTTCAGAACAACTTGGATGAACCTTTTCATGCTCATGCAATTTTCTTGATGCGG 1411

Db 2705 GGACACCATTACAAAACAATCTGGAAGAGTTGTTTTCATCTGCTCAACTTTTCTCACCCCGG 2764

QY 1412 GGAAGTTTGAAGTTTGAGGAGGTTTCCAGGAGGAGTTCAAAAGATATTAAATCAAGAGGAGC 1471

Db 2765 AGAGGTTCCACAATTTGGAAGGTTTTTTGGAGGAGTTTGTGACATTGCCAAGGAGGACC 2824

QY 1472 AGATCTCAAGGTTGCAAAAAATGTTGGCTCCACATTTGCTCAGAAAGSGTAAAAAAGACG 1531

Db 2825 AGATAAAAAAACTGCATGACATGCTGGGGCCGCACATGTTGCGCGGCTCAAAGCCGATG 2884

QY 1532 TAATGAAAGACATGCCCCCAAAAAAGGAGCTCATTTTGGCGTTGATCTGAGCAGTCTGC 1591

Db 2885 TGTTCGAAGAACATGCCCTCCAAGACAGAACTAATTGTGCGTGTGGAGCTGAGCCCTATGC 2944

QY 1592 AGAAAGAAATATTACAAAGTATTTTTTACCCGTAATATCAAGTATTGA---CAAAAAAGG 1648

Db 2945 AGAAGAAATACTACAAGTACATCCTCACTCGAAATTTTGAAGCACTCAATGCCCGAGGTG 3004

QY 1649 GAGGTGCTCAAAATTTCCCTTAATAACATTTATGATGGAATTACGAAAAAGTATGCTGCCATC 1708

Db 3005 GTGGCAACCAGGTGTCCTGCTGAATGTGGTATGATGATCTTAAAGAAGTCTGCAACCATC 3064

QY 1709 CTTA-----TATGCTAGAGGGTGTGTAGCCAGTTATTCACGACGCAAAATGAAGCTT 1759

Db 3065 CATACCTCTTCCCTGTGGCTGCAATGGAAGCTCCTAAGATGCCTAATGSCATGTATGATG 3124

QY 1760 TCAAAACAACCTTTTGGAGTCTTGTGGAAGCTGCAACTTCTAGATAAAATGATGGTCAAAC 1819

Db 3125 GCAGTGCCCTAATCAGAGCATCTGGGAAATTATTGCTGCTGCAGAAATGCTCAAGAAC 3184

QY 1820 TGAAGAGCAAGGACACAGAGTCCTAATATACACACAGTTTTCAGCATATGCTGGACTTAC 1879

Db 3185 TTAAGGAGGTGGGCATCGTGTACTCATCTTTTCCAGATGACCAAGATGCTAGACCTGC 3244

QY 1880 TTGAAGACTACTGTACCCATAAGAAATGGCAGTACGAGCGAAATTGATGGAAGGTTGGCG 1939

Db 3245 TAGAGGATTTCTTGGAAACATGAAGGTTATAAATACGAACGCATCGATGGTGAATCACTG 3304

QY 1940 GAGCTGAGCGGCAAAATACGCATAGATCGGTTCAATGCAAAAAAATCTAACAAAGTTTGT 1999

Db 3305 GGAACATGCGGCAAGAGGCCATTGACCGCTTCAATGCAACCGGCTGCTCAGCAGTTCTGCT 3364

QY 2000 TTTTGTCTCCACAAGAGCTGGTGGCTTAGGAATAAATCTTGCAACGGCTGATACAGTAA 2059

Db 3365 TCTTGCTTCCACTCGAGCTGGGGCCCTTGAATCAATCTGCCACTGTCTGACACAGTTA 3424

QY 2060 TCATTTATGACAGTGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC 2119

Db 3425 TTATCTATGACTCTGACTGGAACCCCAATAATGACATTCAGGCCCTTTAGCAGAGCTCACC 3484

QY 2120 GACTTGGCCAAAACAAATAAGGTGATGATTTATAGGCTCATAAACCGAGGCACCATGAAG 2179

Db 3485 GGATGGGCAAAATAAAAAGGTAATGATCTACCGGTTTGTGACCCCGTGCCTCAGTGGAGG 3544

QY 2180 AAAGGATGATGCAATTGACTFAAAAAAGAAAATGGTTCTAGAGCATCTTGTGTTGG----- 2234

Db 3545 AGCGCATCACGAGGTGGCAAAAGAAAGAAAATGATGCTGACGCATCTAGTGTGGGSCCTG 3604

QY 2235 ----GAAACTCAAAAACAAAAACATTAAATCAGGAAGAGTTAGATGACATCATCAGGTATG 2290

Db 3605 GGCTGGGCTCCAAGACTGGATCTATGTGTCCAAACAGGAGCTTGTATGATATCCTCAAATTTG 3664

QY 2291 GATCAAGGAGCTTTTGTCTAGTGAAGATGATGAAGCAGGAAGTCTTGGAAAAAATTCATT 2350

Db 3665 GCACTGAGGAACCTATTCAAGGATGAAGCCACTGATGGAGGAGAGACAACAAGAGGGAG 3724

QY 2351 ATGATGATGCGGCTATAGACAAAATTGCTTGA-----TCGTGATCTCGTGGAGGCAGAGG 2404

Db 3725 AAGATAGCAGTGTATTCCACTACGATGATAAGGCCATTGAACGGCTGTAGACCGTAACC 3784

QY 2405 AAGTCTCAGTGGATGATGAAGAGGAGAATGGATTCTTAAAGGCTTTCAAGGTGGCTAATT 2464

Db 3785 AGGATGAGACTGAAGACACAGAATTGCAGGGCATGAATGAATATTGAGCTCATTCAAAG 3844

QY 2465 TTGAATATATAGATGAAAAATGAGSCAGCAGCATTTAGAGSCACAGAGAGTCGCTGCTGAAA 2524

Db 3845 TGGCCCCAGTATGTGGTACGGGAAGAAATAATGGGGAGGAAGAGGAGGTAGAACGGGAAA 3904

QY 2525 GCAAACTTTCAGCAGGCAATTCTGATAGACAAAGTTATTGGGAAGAGTTGTTAAAGATA 2584

Db 3905 TCATTAAACAGGAAGAAAGTGTGGATCCTGA---CTACTGGGAGAAAATTGCTGCGGCACC 3961

QY 2585 AATTTAGCTGCACCAGGCTGAGGAGCTTAATGCTCTTGTGTTTGGAAAGATGTGAGCTCTG 2644

Db 3962 ATTATGAGCAGCAGCAAGAAGATCTAGCCCCGAAATCTGGGCAAGGAAAAAAGAATCCGTA 4021

QY 2645 AGCAGTTGGTATCCATTGAAG---AAGATGATCTTGTGTTTGGAAAGATGTGAGCTCTG 2701

Db 4022 AACAGGTCAACTACAATGATGGCTCCCAGGAGGACCCGAGATTGGCAGGACGACCAGTCCG 4081

QY 2702 ATGGAGATGAAAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGACAAAGGAGTTC 2761

Db 4082 ACAACCACTCCGATTACTCAGTGGCTTCAGAGGAAGGTGATGAAGACTTTGATGAACGTT 4141

QY 2762 AGACGGGTCGACGGCCGTACAGAAAGGAAGGGTCGCGATAAATTTGGAACCAA----- 2812

Db 4142 CAGAAGCTCCCGTAGGCCCAGTCGTAAGGCCCTGCGGAATGATAAAGATAAGCCATTGC 4201

QY 2813 CTCGGTTGATGGAAGGTGAGGGGAGATCTTTTCAGAGTACTGGGTTTCAACCAAGAGTCAAA 2872

Db 4202 CTCCTCTGTTGGCCCCGTGTTGGTGGGAATATTGAAGTACTTGGTTTAAATGCTCGTCAGC 4261

QY 2873 GGGCCATTTTGTACAGACTTTGATGAGGTATGG-----AGCTGGCAATTTTGATT 2923

Db 4262 GAAAAGCCTTTCTTAATGCAATTATGCGATATGGTATGCCACCTCAGGATGCTTTTACTA 4321

QY 2924 GGAAGGAGTTTGTTCCTCGCTTAAAGCAGAAAGACCTTTTGAAGAAATAAATGAATATGGAA 2983

Db 4322 CCCAGTGGCTTGTAAAGAGACCCTGCGAGGCAAAATCAGAGAAAGAGTTCAAGGCATATGTCT 4381

QY 2984 TACTCTTCTTGAAGCACATTGCTGAAGAAATAGACGAGAAATTCCTCCAACCTTTTTCAGATG 3043

Db 4382 CTCTTTTCATGCGGCATTTATGTGAGCCGGGGCAGATGGGGCTGAGACCTTTGCTGATG 4441

QY 3044 GTGTGCCCCAAGGAAGGACTTAGAATAGAGATGTTCTAGTCAGAAATGCTTCTCTGATAC 3103

Db 4442 GTGTCCCCCGAAGGCCCTGTCTCGCCACGATGTCTTACTAGAATGGTGTATGTCTT 4501

QY 3104 TAGTTCAGGAAGGTGAAATTTGTAGAAGATCATCCAGGG 3144

Db 4502 TGATTGCAAGAAGGTTTCAGGAGTTTGAACATGTTAATGGG 4542

RESULT 2  
US-09-249-181A-1  
; Sequence 1, Application US/09249181A  
; Patent No. 6440679  
; GENERAL INFORMATION:  
; APPLICANT: Seelig, Hans Peter  
; APPLICANT: Renz, Manfred

; TITLE OF INVENTION: DERMATOMYOSITIS-SPECIFIC AUTO-ANTIGEN
; FILE REFERENCE: 8484-0059-999
; CURRENT APPLICATION NUMBER: US/09/249,181A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 08/913,832
; PRIOR FILING DATE: 1998-01-12
; PRIOR APPLICATION NUMBER: PCT/DE96/00444
; PRIOR FILING DATE: 1996-03-08
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 6328
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(5736)
US-09-249-181A-1

Query Match 10.1%; Score 421.4; DB 4; Length 6328;
Best Local Similarity 51.9%; Pred. No. 1.2e-111;
Matches 1236; Conservative 0; Mismatches 1076; Indels 69; Gaps 10;

QY 818 TACATCCATACCAAGCTTGAGGGACTTAATTTTTCGCGTTCCTCGTGGTCAAAACAGACGC 877
Db 2177 TGCACCCCTATCAAAATGGAGGGCCTGAATGTGTGCGCTTCTCCTGGCTCAGGGCACTG 2236

QY 878 ATGTAATCCTTCGTGATGAAATGGGACTAGGCAAGACAAATTCAAAGCATTCGCCCTTTTAG 937
Db 2237 ACACCATCTTGGCTGATGAGATGGGCCCTTGGGAAACTGTACAGACAGCAGTCTTCCTGT 2296

QY 938 CTTCACCTTTTGGAGG-----AACCTCATTCGGCATTTGGTAATTCCTCCTCTATCGA 991
Db 2297 ATTCCCTTTACAGGAGGGTCAATTCAAAAGGCCCTTCTCTAGTGAGCGCCCTCTTTCTA 2356

QY 992 CTCCTGCGTAACCTGGGAGAGAGAGTTTGGCCATGTCGAGCAGTTATFACAGAGAACATGAGTTTACTTATCGAAAG 1051
Db 2357 CCATCATCAACTGGAGCGGGAGTTTGAATGTGGGCTCCAGACATGTATGTCGTAACCT 2416

QY 1052 ATTTTGGCACTGCGCAAGCTCGAGCAGTTATFACAGAGAACATGAGTTTACTTATCGAAAG 1111
Db 2417 ATGTGGGTGACAAAGGACAGCCGTCGCATCATCCGAGAGAATGAGTTCTCTTTGAAGACA 2476

QY 1112 ATCAAAAAAAGATCAAGAAAAAGAAATCTGGACAAATAAGTAGCGGAAAGCAAGCAAAAA 1171
Db 2477 ATGCCATTCGTGGTGCAAGAGGCCTC-----CCGCATGAAGAAAGAGGCAT 2524

QY 1172 GAATCAAGTTTGATGCTCCTCCTCACATCGTATGAGATGATCAACCTAGATTCAGCAGTTC 1231
Db 2525 CTGTGAAATTCATGTGCTGTGACATCCTATGAAATTGATCACCATTGACATGGCTATTT 2584

QY 1232 TAAACCAATTAAGTGGGAGTGCATGATTGTTGATGAAGGTTCATCGACTGAAAAATAAGG 1291
Db 2585 TGGGCTCTATTGATTTGGGCTGGCCTCATCGTGGATGAAGCCCATCGGCTGAAGAACAATC 2644

QY 1292 ATTCAAAAGCTGTTCTCTTCATTGACACAGTATTCAAAGTAACCCCGTATTCCTTGACAG 1351
Db 2645 AGTCTAAGTTCTTCGGGTATTGAATGGTTACTCACTCCAGCACAAAGCTGTTGCTGACTG 2704

QY 1352 GAACACCACCTTCAGAACAACTTGGATGAACITTTTCATGCTCATGCTATTTCTTGATCGCG 1411
Db 2705 GGACACCATTACAAAAACAATCTGGAAGAGTTGTTTCATCTGCTCAACTTCTCACCCCGG 2764

QY 1412 GGAAGTTTGGAGTTTGGAGGAGTTCAGGAGGAGTTCAAAGATATTAATCAAGAGGAGC 1471
Db 2765 AGAGGTTCCACAATTTGGAAGGTTTTCGAGGAGTTTGTGACATTCGCAAGGAGGACC 2824

QY 1472 AGATCTCAAGTTTGCAAAAAATGTTGGCTCCACATTTGCTCAGAAAGGGTAAAAAAGACG 1531
Db 2825 AGATAAAAAAACTGCATGACATGCTGGGCGCGACATGTTGCGCGGCTCAAAGCCGATG 2884

QY 1532 TAATGAAAGACATGCCCCCAAAAAGGAGCTCATTTTTCGTTGTTGATCTGAGCAGTCTGC 1591

Db 2885 TGTTCAAGAAACATGCCCTCCACAGACAGAACTAATTGTGCGTGTGAGCTGAGCCCTATGC 2944
QY 1592 AGAAAGAATATTACAAAGCTAATTTTACCCGTAATTAATCAAGTATTGA---CAAAAAAGG 1648
Db 2945 AGAAGAAATACTACAAGTACATCCTCACTCGAAAATTTTGAACACACTCAATGCCGAGGTG 3004
QY 1649 GAGGTGCTCAAAATTTCCCTTAATAACATTAATGATGGAATTACGAAAAAGTATGCTGCCATC 1708
Db 3005 GTGGCAACCAGGTGTCTCTGCTGAATGTGGTATGATGATCTTAAGAAGTGTGCAACCATC 3064
QY 1709 CTTA-----TATGCTAGAGGGTGTGAGCCAGTTATTACAGACGCAAAATGAAGCTT 1759
Db 3065 CATACCTCTTCCCTGTGGCTGCAATGGAAGCTCCTAAGATGCTAATGGCATGTATGATG 3124
QY 1760 TCAAAACAACCTTTTGGAGTCTTGTGGAAAAGCTGCAACTTCTAGATAAAATGATGTCAAAC 1819
Db 3125 GCAGTGGCCTAATCAGAGCATCTGGGAAAATTAATTGCTGCTGCAGAAAAATGCTCAAGAACC 3184
QY 1820 TGAAGAGCAAGGACACAGAGTCTCTAATAATATACACACAGTTTTCAGCATATGCTGGACTTAC 1879
Db 3185 TTAAGGAGGTGGGCATCGTGTAATCTCTTTCCAGATGACCAAGATGCTAGACCTGC 3244
QY 1880 TTGAAGACTACTGTACCCCATAGAANAATGGCAGTACGAGCGGAATTGATGGAAGGTTGGCG 1939
Db 3245 TAGAGGATTTCTTGGAAACATGAAGGTTATAAATACGAACGCATCGATGGTGAATCACTG 3304
QY 1940 GAGCTGAGCGGCAAAATACGATAGATCGGTTCAATGCCAAAAATTTCTAACAAGTTTTGTT 1999
Db 3305 GGAACATGCGGCAAGAGGCCATTGACCGCTTCAATGCACCGGTGCTCAGCAGTTCTGCT 3364
QY 2000 TTTTGTCTCCACAAGAGCTGTGGCTTAGGAATAAAATCTTGCAAACGGCTGATACAGTAA 2059
Db 3365 TCTTGTCTTCCACTCGAGCTGGGGCCTTGAATCAATCTGGCCACTGCTGACACAGTTA 3424
QY 2060 TCATTTATGACAGTGACTGGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC 2119
Db 3425 TTATCTATGACTCTGACTGGAAACCCCATATAATGACATTCAGSCCTTTAGCAGAGCTCACC 3484
QY 2120 GACTTGGCCAAACAAATAAGGTGATGATTTATAGGCTCATAAACCGAGGCCCATTTGAAG 2179
Db 3485 GGATTGGGCAAAATAAAAGGTAATGATCTACCGTTTTGTGACCCGTGCGTCAGTGGAGG 3544
QY 2180 AAAGATGATGCAATTGACTAAAAAGAAAATGGTTCTAGAGCATCTTGTGTTGG----- 2234
Db 3545 AGCGCATCACGAGGTGGCAAGAAAGAAAATGATGCTACGCATCTAGTGTGCGGCCTG 3604
QY 2235 ----GAAACTCAAAACACAAAACATTAATCAGGAAGAGTTAGATGACATCATCAGGTATG 2290
Db 3605 GGCTGGGCTCCAAGACTGGATCTATGTCCAACAGGAGCTTGATGATATCCTCNAATTTG 3664
QY 2291 GATCAAGGAGCTTTTGTCTAGTGAAGATGATGAAGCAGGAAAAGTCTGAAAAAATTCATT 2350
Db 3665 GCCTGAGGAACATTTCAAGGATGAAGCCACTGATGGAGGAGAGACAAAGAGGGAG 3724
QY 2351 ATGATGATCGGGCTATAGACAAATTGCTTGA-----TCGTGATCTCGTGGAGGCAGAGG 2404
Db 3725 AAGATAGCAGTGTATTCCACTACGATGATAAGGCCATTGAACGGCTGCTAGACCGTAACC 3784
QY 2405 AAGTCTCAGTGGATGATGAAGAGGAGAAATGGATTCTTTAAAGCTTTCAAGGTGGCTAATT 2464
Db 3785 AGGATGAGACTGAAGACACAGAAATTCAGGGCATGAATGAATTTGAGCTCATTTCAAAG 3844
QY 2465 TTGAATATATAGATGAAAAATAGGCAGCAGCATATTAGAGGCACAGAGAGTTCGTGTGAAA 2524
Db 3845 TGGCCAGTATGTGGTACGGGAAGAAAGAAATGGGGAGGAAGAGGAGGTAGAACGGGAAA 3904
QY 2525 GCAATCTTCAGCAGGCAATTCGTATAGAGCAAGTTATTGGGAAGAGTTGTTAAAAAGATA 2584
Db 3905 TCATTAAACAGGAAGAAAGTGTGATCCTGA---CTACTGGGAGAAATGCTGCGGCACC 3961
QY 2585 AATTTGAGCTGCACCAGGCTGAGGAGCTTAATGCTCTTGGAAAAAGGAAGAGATCGCA 2644
Db 3962 ATTATGAGCAGCAGCAAGAGATCTAGCCCCGAAATCTGGGCAAGGAAAAAGAAATCCGTA 4021





Db 3365 TCTTGCTTCCACTCGAGCTGGGGCCCTTGGAAATCAATCTGGCCACTGCTGACACAGTTA 3424

Qy 2060 TCATTTATGACAGTGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC 2119

Db 3425 TTATCTATGACTGTAACCCCAATATGACATTCAGGCCCTTAGCAGAGCTCACC 3484

Qy 2120 GACTTGGCCAAACAAATAAGGTGATGATTTATAGGCTCATAAACCGAGCCACCATTTGAAG 2179

Db 3485 GGATTGGGCAATAAAGGTAAATGATCTACCGGTTTGTACCCGTCGCTCAGTGGAGG 3544

Qy 2180 AAAGGATGATGCAATTGACTAAAAAGAAATGGTTCTAGAGCATCTTGTGTTGTTGG----- 2234

Db 3545 AGCGCATCACGCAAGTGGCAAGAAAGAAATGATGCTGACGCACTAGTGGTGGCCTG 3604

Qy 2235 ---GAAACTCAAAACACAAAACATTAATCAGGAAGAGTTAGATGACATCATCAGGTATG 2290

Db 3605 GGCTGGGCTCCAAGACTGGATCTATGTCCAAACAGGAGCTTGTATGATATCCTCAAAATTG 3664

Qy 2291 GATCAAAAGGAGCTTTTGTAGTGAAGATGATGAAGCAGGAAAGTCTGGAATAATTCATT 2350

Db 3665 GCACTGAGGAACATATTCAAGGATGAAGCCACTGATGGAGGAGAGACAACAAGAGGGAG 3724

Qy 2351 ATGATGATCGGCTATAGACAAAATTGCTTGA-----TCGTGATCTCGTGGAGGCAGAGG 2404

Db 3725 AAGATAGCAGTGTATCCACTACGATGATAAGGCCATTGAACGGCTGCTAGACCGTAACC 3784

Qy 2405 AAGTCTCAGTGATGATGAAGAGGAGATGGAATCTTAAAGGCTTCAAGGTGGCTAATT 2464

Db 3785 AGGATGAGACTGAAGACACAGAATTGCAAGGCATGAATGAATATTGAGCTCATTCAAAG 3844

Qy 2465 TTGAATATATAGTGAATAATGAGGCAGCAGCATTAGAGGCACAGAGTCTGCTGAAA 2524

Db 3845 TGGCCCAAGTATGTGTTACGGGAAGAAAGAAATGGGGGAGGAAGAGGTAGAACGGGAAA 3904

Qy 2525 GCAAATCTTCAGCAGGCAATTCTGATAGAGCAAGTTATTGGGAAGAGTTGTTAAAAGATA 2584

Db 3905 TCATTAAACAGGAAGAAAGTGTGGATCTCTGA---CTACTGGGAGAAATGCTGCGGCACC 3961

Qy 2585 AATTTGAGCTGCACCAGGCTGAGGAGCTTAATGCTCTTGGAAAAAGGAAGAGAGTCGCA 2644

Db 3962 ATTATGAGCAGCAGCAAGAAGATCTAGCCCCGAAATCTGGGCAAGGAAGAAAGAAATCCGTA 4021

Qy 2645 AGCAGTTGGTATCCATTGAAG--AAGATGATCTTGTGTTTGGAAAGATGTGAGCTCTG 2701

Db 4022 AACAGGTCAACTACAATGATGGCTCCAGGAGGACCGAGATTGGCAGACGACCAAGTCCG 4081

Qy 2702 ATGGAGATGAAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGACAAGGAGTTC 2761

Db 4082 ACAACCACTCCGATTAATCAGTGGCTTACAGAGGAAGGTGATGAAGACTTTGATGAACGTT 4141

Qy 2762 AGACGGGTGACGCGGTACAGAAGAAAGGTGCGGATAAATTTGGAAACCAA----- 2812

Db 4142 CAGAAGCTCCCCGTAGGCCCAAGTCGTGAAGGCTCGGAATATGATAAAGATAAGCCATTGC 4201

Qy 2813 CTCCGTTGATGGAAGGTGAGGGGAGATCTTTCAGAGTACTGGGTTTCAACACAGAGTCAAA 2872

Db 4202 CTCCTCTGTTGGCCCGTGTGGTGGGAATATGAAGTACTTGGTTTAAATGCTCGTCAGC 4261

Qy 2873 GGGCCATTTTGTACAGACTTTTGATGAGGTAATGG-----AGCTGGCAATTTTGATT 2923

Db 4262 GAAAAGCCTTTCTTAATGCAATTAATGCGATATGGTATGCCACCTCAGGATGCTTTTACTA 4321

Qy 2924 GGAAGGAGTTTGTTCCTCGCTTAAAGCAGAGAAGCCTTTGAAGAAATAAATGAATATGGAA 2983

Db 4322 CCCAGTGGCTTGAAGAGACCTGCGAGGCAATCAGAGAAAGAGTTCAAGGCATATGTCT 4381

Qy 2984 TACTCTTCTTGAAGCACATTTGCTGAAGAAATAGACGAGAAATCTCCAACTTTTCAGATG 3043

Db 4382 CTCCTTTTCATCGCGCATTTATGTGAGCCGGGGCAGATGGGGCTGAGACCTTTGCTGATG 4441

Qy 3044 GTGTGCCCCAAGGAGCACTTAGAATAGAAGATGTTCTTAGTCAGAAATTGCTCTTCTGATAC 3103

Db 4442 GTGTCCCCCGAAGAGCCCTGTCTCGCCAGCATGTCCTTACTAGAAATGGTGTATGTCTT 4501

Qy 3104 TAGTTCAGGAGAAAGGTGAATTTGTAGAAGATCATCCAGGG 3144

Db 4502 TGATTCGAAGAAGGTTTCAGGAGTTTGAACATGTTAATGGG 4542

RESULT 4

US-09-620-312D-325  
; Sequence 325, Application US/09620312D  
; Patent No. 6569662  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yunqing  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: John Tillinghast  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; CURRENT FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: pt\_FL\_genes Version 1.0  
; SEQ ID NO 325  
; LENGTH: 6475  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (219)..(5957)  
US-09-620-312D-325

Query Match 10.1%; Score 421.4; DB 4; Length 6475;  
Best Local Similarity 51.9%; Pred. No. 1.2e-111;  
Matches 1236; Conservative 0; Mismatches 1076; Indels 69; Gaps 10;

Qy 818 TACATCCATACCAGCTTGAGGGACTTAATTTTTTTCGGTTCTCGTGGTCAAAAACAGACGC 877

Db 2395 TGCACCCCTATCAAAATGGAGGSCCTGAATTGGTTGCGCTTCTCCTGGGCTCAGGGCACTG 2454

Qy 878 ATGTAATCCTTGCTGATGAATGGGACTAGGCAAGACAATTCAAAGCATTTGCCCTTTTAG 937

Db 2455 ACACCATCTTGGCTGATGAGATGGGCCCTTGGGAAAACCTGTACAGACAGCAGTCTTCTCTGT 2514

Qy 938 CTTCACTTTTGGAGGAG-----AACCTCATTCGCGCATTTGGTAATTGCTCCTCTATCGA 991

Db 2515 ATTCCCTTTACAAGGAGGGTCAATCCAAAGGCCCTTCTCCTAGTGAGCGCCCTCTTTCTA 2574

Qy 992 CTCTGCGTAACTGGGAGAGAGAGTTTGCCACATGGGCCCCACAGATGAACGTTGTTATGT 1051

Db 2575 CCATCATCAACTGGGAGCGGGAGTTTGAATGTGGGCTCCAGACATGTATGTCGTAACCT 2634

Qy 1052 ATTTTGGCACTGCGCAAGCTCGAGCAGTTTATCAGAGAACATGAGTTTACTTATCGAAAG 1111

Db 2635 ATGTGGGTGACAAGGACAGCCGTCGCATCATCCGAGAGATGAGTTCTCCTTTGAAGACA 2694

Qy 1112 ATCAAAAAAAGATCAAGAAAAAAGAAATCTGGACAAATAAGTAGCGAAAGCAAGCAAAAAA 1171

Db	2695	ATGCCATT	CGTGGTGGCAAGGCCTC-----CCGCATGAAGAAAGAGGCAT	2742	
QY	1172	GAATCAAGTTT	TGATGTCCTCCTCACATCGTATGAGATGATCAACCTAGATTACAGCAGTTTC	1231	
Db	2743	CTGTGAAATTC	CATGTGCTGTGACATCCTATGAATTGATCACCATTGACATGGCTATTT	2802	
QY	1232	TAAACCAATTA	AGTGGAGTGTCATGATTGTTGATGAAGGTTCATCGACTGAAAAATAAAGG	1291	
Db	2803	TGGGCTCTATT	TGATTGGCCTGCCTCATCGTGGATGAAGCCCATCGGCTGAAGAAACAATC	2862	
QY	1292	ATTCAAAGCTGT	CTCTTCATTGACACAGTATTCAAGTAAACCACCGTATTCTTCTGACAG	1351	
Db	2863	AGTCTAAGTTCT	TCGGGTATTGAATGGTTACTCACTCCAGCACAAAGCTGTTGCTGACTG	2922	
QY	1352	GAAACACCAC	CTTCAGAACTTGGATGAACCTTTTCATGCTCATGCAATTTCTTGATGCGG	1411	
Db	2923	GGACACCATT	TACAAAACAATCTGGAAGAGTTGTTTCATCTGCTCAACTTCTCACCCCG	2982	
QY	1412	GGAAGTTTGA	AGTTTGGAGGAGTTCCAGGAGGAGTTCAAAAGATATTAAATCAAGAGGAGC	1471	
Db	2983	AGAGGTTCC	ACAATTTGGAAGGTTTTTTGGAGGAGTTTGCTGACATTGCCAAGGAGGACC	3042	
QY	1472	AGATCTCAAG	GTTCACAAAATGTTGGCTCCACATTTGCTCAGAAGGTAAAAAAGACG	1531	
Db	3043	AGATAAAAAA	AACTGCATGACATGCTGGGGCCGACATGTTGCGGGCTCAAAGCCGATG	3102	
QY	1532	TAATGAAAG	ACATGCCCCCAAAAAGGAGCTCATTTTGCCTGTTGATCTGAGCAGTCTGC	1591	
Db	3103	TGTTCAAGA	ACATGCCCTCCAAGACAGAACTAATTGTGCGTGTGGAGCTGAGCCCTATGC	3162	
QY	1592	AGAAAGAAT	ATTACAAAGCTATTTTACCCGTAATTATCAAGTATTGA---CAAAAAAGG	1648	
Db	3163	AGAAGAAAT	ACTACAAGTACATCCTCACTCGAAATTTTGAAGCACTCAATGCCCGAGGTG	3222	
QY	1649	GAGGTGCTC	AAATTTCCCTTAATAACATTTATGATGGAATTACGAAAAGTATGCTGCCATC	1708	
Db	3223	GTGGCAACC	AGGTGCTCTGCTGAAATGTGGTGATGGATCTTAAGAAGTGCTGCACCATC	3282	
QY	1709	CTTA-----	TATGCTAGAGGGTGTGTAGCCAGTTATTTCACGACGCAATGAAGCTT	1759	
Db	3283	CATACCTCT	TTCCCTGTGGTGCAAATGGAAGCTCCTAAGATGCCCTAATGGCATGTATGATG	3342	
QY	1760	TCAAAACA	ACTTTTGGAGTCTTGTGAAAAGCTGCAACTTCTAGATAAAAATGATGTCAAAAC	1819	
Db	3343	GCAGTGCC	CTAATCAGAGCATCTGGGAAATTAATTGCTGTGCAGAAAAATGCTCAAGAAC	3402	
QY	1820	TGAAAGAG	CAAGGACACAGAGTCTTAATATACACACAGTTTCAGCATATGCTGGACTTAC	1879	
Db	3403	TTAAGG	AGGGTGGGCATCGTGTA	CTCTTTTCCCAGATGACCAAGATGCTAGACCTGC	3462
QY	1880	TTGAAG	ACTACTGTACCAATAAGAAATGGCAGTACGAGCGAATTGATGGAAAGGTTGGCG	1939	
Db	3463	TAGAGG	ATTCTTGGAAACATGAAGGTTATAAAATACGAACGCATCGATGGTGAATCACTG	3522	
QY	1940	GAGCTG	AGCGGCAAAATACGCATAGATCGGTTCAATGCCAAAAAATCTAACAAGTTTGT	1999	
Db	3523	GGAACAT	GCGGCAAGAGGCCATTGACCGCTTCAATGCACCGGGTGTCTCAGCAGTTCTGCT	3582	
QY	2000	TTTTGCT	CTCCACAAGAGCTGGTGGCTTAGGAATAAATCTTGCAACGGCTGATACAGTAA	2059	
Db	3583	TCTTGCT	TTCCACTCGAGCTGGGGCCCTTGGAAATCAATCTGGCCACTGCTGACACAGTTA	3642	
QY	2060	TCATTT	TATGACAGTGGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC	2119	
Db	3643	TTATCT	ATGACTCTGACTGGAACCCCAATAATGACATTCAGGCCCTTTAGCAGAGCTCACC	3702	
QY	2120	GACTTGG	CCAAAACAATAAGGTGATGATTTATAGGCTCATAAAACCGAGGCACCATTTGAAG	2179	
Db	3703	GGATTGG	CAAAATAAAAAGTAATGATCTACCGGTTTGTGACCCCGTGCGTCAGTGGAGG	3762	
QY	2180	AAAGG	ATGATGCAATTGACTAAAAAGAAATGGTTCTTACAGCATCTTGTGTTGG-----	2234	

Db	3763	AGCGCATCAGCAGGTGGCAAAAGAAATAATGATCTGACGCATCTAGTGGTGGCGCTG	3822
Qy	2235	----GAAACTCAAAAAACAAAAACATTAAATCAGGAAGAGTTAGATGACATCATCAGGTATG	2290
Db	3823	GGCTGGGCTCCAAGACTGGATCTATGTCCAAACAGGAGCTTGATGATATCCTCAAATTTG	3882
Qy	2291	GATCAAAAGGAGCTTTTGTAGTGAAGATGATGAAGCAGGAAAGTCTGGAAAAATTCATT	2350
Db	3883	GCACTGAGGAACCTATTCAAGGATGAAGCCACTGATGGAGGAGGAGACAAACAAGAGGGAG	3942
Qy	2351	ATGATGATGCGGCTATAGACAAATTGCTTGA-----TCGTGATCTCGTGGAGGCAGAGG	2404
Db	3943	AAGATAGCAGTGTATTCCACTACGATGATAAGGCCATTGAACGGCTGCTAGACCGTAACC	4002
Qy	2405	AAGTCTCAGTGGATGATGAAGAGGAGGAATGGATTCTTAAAGGCTTTCAGGTGGCTAATT	2464
Db	4003	AGGATGAGACTGAAGACACAGAATTGCAGGCGCATGAATGAATATTTAGCTCATTCAAAG	4062
Qy	2465	TTGAATATATAGATGAAAAATGAGGCAGCAGCATTTAGAGGCACAGAGAGTCGCTGTGAAA	2524
Db	4063	TGGCCCACTATGTGTACGGGAAGAAAGAAATGGGGAGGAAGAGGAGGTAGAACGGGAAA	4122
Qy	2525	GCAAATCTTCAGCAGGCAATTCTGATAGAGCAAGTTATTGGGAAGAGTTCTTAAAAAGATA	2584
Db	4123	TCATTAAACAGGAAGAAAGTGTGGATCCTGA---CTACTGGGAGAAATTCCTGCGGCACC	4179
Qy	2585	AATTTGAGCTGCACCAGGCTGAGGAGCTTAATGCTCTTGGAAAAAGGAAGAGAAAGTCGCA	2644
Db	4180	ATTATGAGCAGCAGCAAGAAGATCTAGCCGAAATCTGGGCAAGGAAAAAGAAATCCGTA	4239
Qy	2645	AGCAGTTGGTATCCATTGAAG--AAGATGATCTTGTGTTTGGAAAGATGTGAGCTCTG	2701
Db	4240	AACAGGTCAACTACAATGATGGCTCCAGGAGGACCGAGATTGGCAGGACGACCAAGTCCG	4299
Qy	2702	ATGGAGATGAAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGACAAGGAGTTC	2761
Db	4300	ACAACCAGTCCGATTACTCAGTGGCTTCAGAGGAAGGTGATGAAGACTTTTGATGAACGTT	4359
Qy	2762	AGACGGTGCAGCGCCGTACAGAAGAAAGGTCGCGATAATTTGGAAACCAA-----	2812
Db	4360	CAGAAAGCTCCCGTAGGCCAGTCGTAAAGGCGCTCGGAATGATAAAGATAAGCCATTGC	4419
Qy	2813	CTCCGTTGATGGAAGGTGAGGGGAGATCTTTTCAGAGTACTGGGTTTCAACAGAGTCAAA	2872
Db	4420	CTCCTCTGTTGGCCCGTGTGGTGGGAATATTGAAGTACTTTGGTTTTTAATGCTCGTCAGC	4479
Qy	2873	GGGCCATTTTGTACAGACTTTTGATGAGGTATGG-----AGCTGGCAATTTTGATT	2923
Db	4480	GAAAAGCCTTTCTTAATGCAATTATGCGATATGGTATGCCACCTCAGGATGCTTTTACTA	4539
Qy	2924	GGAAGGAGTTTGTCTCTCGCTTAAAGCAGAGAAGACCTTTGAAGAAAAATAATGAATATGAA	2983
Db	4540	CCCAGTGGCTTCTAAGAGACCTGCGAGGCAAAATCAGAGAAAGAGTTCAAGSCATATGTCT	4599
Qy	2984	TACTCTTCTTGAAGCACATTGCTGAAGAAATAGACGAGAATTCTCCAACCTTTTCAGATG	3043
Db	4600	CTCTTTTCATCGCGCATTTATGTGAGCCCGGGGCGAGATGGGGCTGAGACCTTTGCTGATG	4659
Qy	3044	GTGTGCCCCAAGGAAGCACTTAGAATAGAACATGTTCTTAGTCAGAATTTGCTTCTGTATAC	3103
Db	4660	GTGTCCCCCGAAGAGGCCCTGTCTCGCCAGCATGTCCTTACTAGAAATTTGGTGTATGTCTT	4719
Qy	3104	TAGTTTCAGGAGAAGGTGAAATTTGTAGAAAGATCATCCAGGG	3144
Db	4720	TGATTCGCAAGAAGGTTTCAGGAGTTTGAACATGTTTAATGGG	4760

RESULT 5  
US-09-976-594-75  
; Sequence 75, Application US/09976594  
; Patent No. 6673549  
; GENERAL INFORMATION:  
; APPLICANT: Furness, Michael

; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 75
; LENGTH: 5837
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 3125723CB1
US-09-976-594-75

Query Match 6.8%; Score 285.4; DB 4; Length 5837;
Best Local Similarity 53.5%; Pred. No. 4.3e-72;
Matches 759; Conservative 0; Mismatches 591; Indels 69; Gaps 5;

QY 826 TACCAGCTTGAGGACTTAATTTTTGCGGTTCTCGTGGTCAAAACAGACGCATGTAATC 885
Db 1584 TATCAGCTAGAGGCTAAACTGGCTAGCTCATTCCTGGTGCAAAATAATAGTGAATC 1643
QY 886 CTTGCTGATGAATGGGACTAGGCAAGACAATTCAAAGCATTGCCCTTTTAGCTTCACIT 945
Db 1644 CTTGCTGATGAATGGGCTAGGAAAGACCATCCAGACCATATCATCTCTCCTACCTG 1703
QY 946 TTTGAGGAGAACCTCAT-----TCCGCAATTGGTAATTGCTCCTCTATCGACTCTGCGT 999
Db 1704 TTCCACCAACACAGCTGTATGGCCCTTTCTTTATAGTCGTCCCTTTATCCACCTCAC 1763
QY 1000 AACTGGGAGAGAGATTGGCCACATGGGCCCCACAGATGAACGTGGTTATGTTTGGC 1059
Db 1764 TCATGGCAGAGAGATTGAAATCTGGGCACAGAGATTAACTGAGTGGTTTACATAGG- 1822
QY 1060 ACTGCGCAAGCTCGAGCAGTTATCAGAGAACATGAGTTTACTTTATCGAAAGATCAAAA 1119
Db 1823 -----TGACCTG 1829
QY 1120 AAGATCAAGAAAAAGAAATCTGGACAATAAGTAGCGAAAGCAAGCAAAAAAGAAATCAAG 1179
Db 1830 ATGAGCAGAAATACGATACGGGAATATGAATGGATTCAATCCCAACCAAAAGATTGAAG 1889
QY 1180 TTTGATGTCCTCTCACATCGTATGAGATGATCAACCTAGATTTCAGCAGTTCTAAAAACCA 1239
Db 1890 TTCAACGCACATTATAACAAACATATGAGATGCTCTTTGAAAGATAAGACTGTGCTGGCAGT 1949
QY 1240 ATTAAGTGGAGTGCATGATTGTTGATGAAGGTTCATCGACTGAAATAAAGGATTCAAAG 1299
Db 1950 ATTAACCTGGCCCTTCTGGGAGTGGATGAAGCCCATCGGTTTGAAGAATGATGACTCTTTA 2009
QY 1300 CTGTTCTCTTCAATTGACACAGTATTCAAGTAACCAACCGTATTCTTCTGACAGGAACCA 1359
Db 2010 TTGTATAAAACTCTGATTGATTTCAAAGTCCAACCATAGGCTCCTGATTACGGGGACCCCT 2069
QY 1360 CTTCAGAACCAACTGGATGAACCTTTTTCATGCTCATGCTATTTTCTTGTATGCGGGAGTTT 1419
Db 2070 CTTCAGAAATCCCTCAAAGAGCTCTGGTCTCTGCTGCACTTTATTTATGCGGAGAGTTT 2129
QY 1420 GGAAGTTTGGAGAGTTCCAGGAGGAGTTCAAAGATATTAATCAAGAGGAGCAGATCTCA 1479
Db 2130 GAAATTTGGGAAGATTTTGAAGAAGA--CCATGGGAAGGGGAGAGAGAAATGGTACCAG 2186
QY 1480 AGTTTGACAAAAATGTTGGCTCCACATTTGCTCAGAAAGGTAAAAAAGACGTAATGAAA 1539
Db 2187 AGTCTTCATAAGGTGTAGAGCCTTTCCCTTCTCCGAGAGTCAAAAAAGATGTGGAGAAA 2246
QY 1540 GACATGCCCCCAAAAAAGGAGCTCATTTTGGCTGTTGATCTGAGCAGTCTGCAGAAAGAA 1599
Db 2247 TCCCTTCCTGCTAAAGTGGAACAGATTCTCAGGCTGGAGATGTCAGCCCTTCAGAAACAG 2306

QY 1600 TATTACAAAGCTATTTTACCGTAATTATCAAGTATTGACAAA---AAAGGAGGTGCT 1656
Db 2307 TATTACAAAGTGGATTCTGACCAGGAATTACAAGGCTCTTGCCAAAGGAACAAGAGGCAGC 2366
QY 1657 CAAATTTCCCTTAATAACATATTATGATGGAATTACGAAAAGTATGCTGCCATCCTTATATG 1716
Db 2367 ACATCTGGTTTTCTTAATAATTGTGATGGAACCTGAAAAAATGTTGCAACCACTGCTATCTG 2426
QY 1717 CTAGAGGGTGTGTGAGCCAGTT--ATTACGACGCAAAATGAAGCTTTTCAAAACAACTTTGT 1773
Db 2427 ATTAAACCCCTCTGAAGAAATGAAGGGAAAAATGGACAGGAGATTCTTCTGTCCCTCATA 2486
QY 1774 GAGTCTGTGGAAAGCTGCAACTTCTAGATAAAAATGATGTCAAAACCTGAAAAGAGCAAGGA 1833
Db 2487 AGGAGCAGTGGGAAGTTGATTTTATTATGACAAAACCTGTTGACAAGACTTCGAGAAAAGGGG 2546
QY 1834 CACAGAGTCTTAATATACACACAGATTTCAGCATATGCTGGACTTACTTGAAGACTACTGT 1893
Db 2547 AATCGAGTCTTATCTTCTCTCAGATGGTGAGAATGTTGGATATCTCTGGCTGAATACCTA 2606
QY 1894 ACCATAAGAAATGGCAGTACGAGCGGAATTGATGGAAGGTTGGCGGAGCTGAGCGGCAG 1953
Db 2607 ACTATTAAACACTATCCTTTCCAGCGTCTGGATGGTTCCATCAAGGGAGAAATCCGAAAA 2666
QY 1954 ATACGCATAGATCGGTTCAATGCCAAAAAATTTCTAACAAAGTTTGTGTTTGTCTCTCCACA 2013
Db 2667 CAGGCACTGGACCACCTTCAATGTCAGATGGGTCTGAGGACTTCTGTTTCTGCTCTCGACA 2726
QY 2014 AGAGCTGGTGGCTTAGGAATAAATCTTGCAACGGCTGATACAGTAATCATTTATGACAGT 2073
Db 2727 AGGCTGGTGGCCTGGGAATCAATTTGGCTTCAGCGGACACAGTCGTCTCTTGTGACTCT 2786
QY 2074 GACTGGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGTTCATCGACTTGGCCAAACA 2133
Db 2787 GACTGGAACCCCAAGATGACTTGCAGGCACAAGCCCGAGCGCATAGAAATTGGTCAGAAG 2846
QY 2134 AATAAGGTGATGATTTATAGGCTCATAAACCGAGGCCACCATTTGAAGAAAGGATGATGCAA 2193
Db 2847 AAGCAGGTAAATATTACCGCTTAGTTACAAAGGGGACTGTGGAGGAGGAGATCATAGAA 2906
QY 2194 TTGACTAAAAAGAAAAATGGTTCTTAGAGCATCTTGTGTT 2232
Db 2907 CGGGCCAAAAAGAGATGGTATTAGATCATCTGTGATT 2945

RESULT 6
US-09-702-705-1802
; Sequence 1802, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C14
; CURRENT APPLICATION NUMBER: US/09/702,705
; CURRENT FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 1833
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1802
; LENGTH: 2512
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-702-705-1802



Query Match		5.1%;	Score 213.6;	DB 4;	Length 2512;
Best Local Similarity		53.8%;	Pred. No. 1.9e-51;		
Matches 544;		Conservative 0;	Mismatches 444;	Indels 24;	Gaps 4;
QY	1180	TTTGATGTCCTCCATCGTATGAGATGATCAACCTAGATTTCAGCAGTTCTAAAAACCA	1239		
Db	456	TTTCATGTGCTACTGACTACCTATGAGATTGTTGTTGAAAGATGCATCATTTCTAAATCA	515		
QY	1240	ATTAACTGGGAGTGCATGATTGTTGATGAAGTTCATCGACTGAAAAATAAGGATTCAAAG	1299		
Db	516	TTCCCTTGGAGTGTCTTGTGTGGATGAAGCTACAGGTTGAAAAACCAAGCTCCCTG	575		
QY	1300	CTGTTCTCTTCATTGACACAGTATTCAAAGTAAACCACCGTATTCTTCTGACAGGAACACCA	1359		
Db	576	CTGCATAAGACCTTGTACAGATTCTCAGTAGTCTTCAGTCTCCTGTGACCGGAACCTCCC	635		
QY	1360	CTTCAGAACCACTTTGGATGAACTTTTCATGCTCATGCTATTTCTTGATCGGGGAAGTTT	1419		
Db	636	ATCCAGAACAGCCTCCAAGAGCTCTACTCCCTCCTCAGTTTGTGGAGCCTGATCTCTTT	695		
QY	1420	-----GGAAGTTTGGAGGAGTTCAGGAGGAGTTCAAAAGATATTAAATCAAGAGGACG	1473		
Db	696	TCCAAGGAAGAGGTGGGAGATTTTATTCACCGTACCAGGATATTGAGAAAGAAATCTGAG	755		
QY	1474	ATCTCAAG--GTTGCACAAAATGTTGGTCCACATTTTGCTCAGAAGGGTAAAAAAGAC	1530		
Db	756	TCAGCAAGTGAACCTGCACAAACTCTTCAGGCCATTTCTGCTGAGCGAGTGAAAGCTGAG	815		
QY	1531	GTAATGAAAGACATGCCCCCAAAAGGAGCTCATTTTGCCTGATCTGAGCAGTCTG	1590		
Db	816	GTAGCTACAGAGCTTCCCAAGAAAGACAGAAAGTAGTGATATACCATGGCATGTGAGCATTG	875		
QY	1591	CAGAAAGAAATATCAAAAGCTATTTTACCCTGTAATTATCAAGTATTGACAAAAAGGGA	1650		
Db	876	CAGAAGAAATACTACAAGGCCATTTTGATGAAAGACCTAGATGCATTTGAAAAATGAGACG	935		
QY	1651	GGTGCTCAAAATTTCCCTTAATAACATTTATGATGGAATTACGAAAAAGTATGCTGCCATCCT	1710		
Db	936	GCAAAGAAGGTAAACTACAGAACATTTTGTCCCAGCTTCGAAAGTGTGTGGATCACCCA	995		
QY	1711	TATATGCTAGAGGTTGTTGAGCCAGTTATTACGACCGCAATATTACAGCAACTTTCAAACAAC	1770		
Db	996	TATTTGTTTGAATGGTGTGGAGCC-----GGAGCCTTTTGAAGTTGGAGACCACCTG	1046		
QY	1771	TTGGAGTCTTGTGGAAAGCTGCAACTTCTAGATAAAATGATGTTCAAACTGAAAGAGCAA	1830		
Db	1047	ACTGAGGCTAGTGGGAAGCTTACCTGCTGGATAAGCTACTAGCATTCCTCTGTATTTCTGGG	1106		
QY	1831	GGACACAGAGTCCCTAATATACACACAGTTTTCAGCATATGCTGGACTTACTTGAAGACTAC	1890		
Db	1107	GGCCATCGGGTTTACTTTTCTCCCAAATGACCCAGATGTTGGATATTCTCCAAGACTAT	1166		
QY	1891	TGTACCCATAAGAAATGGCAGTACGAGCGAATTGATGGAAGGTTGGCGGAGCTGAGCGG	1950		
Db	1167	ATGGATTACAGAGGCTACAGCTATGAGCGTGTGGATGGTTCTGTGAGAGGAAGAGAGA	1226		
QY	1951	CAAAATACGCATAGATCGGTTCAATGCCAAAAATTTCTAAACAAGTTTGTGTTTGTGCTCTCC	2010		
Db	1227	CAC TTGGCCATTAA-----GAACTTTGGAACAGCAGCCCATTTTCGTTTTCCTCCTGAGT	1280		
QY	2011	ACAAGAGCTGGTGGCTTAGGAATAAATCTTGCAACCGCTGATACAGTAATCATTTATGAC	2070		
Db	1281	ACTAGGGCAGGTGGAGTTGGCATGAACCTTAACAGCAGCAGATACTGTGATTTTGTGTTGAC	1340		
QY	2071	AGTACTGGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATCGACTTGGCCAA	2130		
Db	1341	AGTGACTTTAATCCTCAGAAATGACTTGTCAAGCAGCTGCCAGGGCTCATCGCATTTGGCCAA	1400		
QY	2131	ACAAATAAGGTGATGATTTATAGGCTCATAAACCAGGCAACCATTTGAAGAA	2182		
Db	1401	AACAAGTCTGTTAAAGTTATTCGGCTGATTGGTCGAGACACTGTGGAAGAA	1452		

RESULT 7	
US-09-736-457-1802	
; Sequence 1802, Application US/09736457	
; Patent No. 6509448	
; GENERAL INFORMATION:	
; APPLICANT: Wang, Tongtong	
; APPLICANT: Bangur, Chaitanya S.	
; APPLICANT: Lodes, Michael A.	
; APPLICANT: Fanger, Gary	
; APPLICANT: Vedvick, Tom	
; APPLICANT: Carter, Darrick	
; APPLICANT: Retter, Marc	
; APPLICANT: Mannion, Jane	
; APPLICANT: Fan, Liqun	
; APPLICANT: Wang, Aijun	
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND	
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER	
; FILE REFERENCE: 210121.478C15	
; CURRENT APPLICATION NUMBER: US/09/736,457	
; CURRENT FILING DATE: 2000-12-13	
; NUMBER OF SEQ ID NOS: 1864	
; SOFTWARE: FastSEQ for Windows Version 3.0	
; SEQ ID NO 1802	
; LENGTH: 2512	
; TYPE: DNA	
; ORGANISM: Homo sapiens	
US-09-736-457-1802	

Query Match		5.1%;	Score 213.6;	DB 4;	Length 2512;
Best Local Similarity		53.8%;	Pred. No. 1.9e-51;		
Matches 544;		Conservative 0;	Mismatches 444;	Indels 24;	Gaps 4;
QY	1180	TTTGATGTCCTCCATCGTATGAGATGATCAACCTAGATTTCAGCAGTTCTAAAAACCA	1239		
Db	456	TTTCATGTGCTACTGACTACCTATGAGATTGCTTGAAGATGCATCATTTCTAAATCA	515		
QY	1240	ATTAACTGGGAGTGCATGATTGTTGATGAAGTTCATCGACTGAAAAATAAGGATTCAAAG	1299		
Db	516	TTCCCTTGGAGTGTCTTGTGTGGATGAAGCTCACAGGTTGAAAAACCAAGCTCCCTG	575		
QY	1300	CTGTTCTCTTCATTGACACAGTATTCAAAGTAAACCACCGTATTCTTCTGACAGGAACACCA	1359		
Db	576	CTGCATAAGACCTTGTACAGATTCTCAGTAGTCTTCAGTCTCCTGTGACCGGAACCTCC	635		
QY	1360	CTTCAGAACCACTTTGGATGAACTTTTCATGCTCATGCTATTTCTTGATCGGGGAAGTTT	1419		
Db	636	ATCCAGAACAGCCTCCAAGAGCTCTACTCCCTCCTCAGTTTGTGGAGCCTGATCTCTTT	695		
QY	1420	-----GGAAGTTTGGAGGAGTTCAGGAGGAGTTCAAAGATATTAAATCAAGAGGACG	1473		
Db	696	TCCAAGGAAGAGGTGGGAGATTTTATTCACCGTACCAGGATATTGAGAAAGAAATCTGAG	755		
QY	1474	ATCTCAAG--GTTGCACAAAATGTTGGTCCACATTTTGCTCAGAAGGGTAAAAAAGAC	1530		
Db	756	TCAGCAAGTGAACCTGCACAAACTCTTCAGGCCATTTCTGCTGAGCGAGTGAAAGCTGAG	815		
QY	1531	GTAATGAAAGACATGCCCCCAAAAGGAGCTCATTTTGCCTGTTGATCTGAGCAGTCTG	1590		
Db	816	GTAGCTACAGAGCTTCCCAAGAAAGACAGAAAGTAGTGATATACCATGGCATGTGAGCATTG	875		
QY	1591	CAGAAAGAAATATCAAAAGCTATTTTACCCTGTAATTATCAAGTATTGACAAAAAGGGA	1650		
Db	876	CAGAAGAAATACTACAAGGCCATTTTGATGAAAGACCTAGATGCATTTGAAAAATGAGACG	935		
QY	1651	GGTGCTCAAAATTTCCCTTAATAACATTTATGATGGAATTACGAAAAAGTATGCTGCCATCCT	1710		
Db	936	GCAAAGAAGGTAAACTACAGAACATTTTGTCCCAGCTTCGAAAGTGTGTGGATCACCCA	995		
QY	1711	TATATGCTAGAGGTTGTTGAGCCAGTTATTACGACCGCAATATTACAGCAACTTTCAAACAAC	1770		
Db	996	TATTTGTTTGAATGGTGTGGAGCC-----GGAGCCTTTTGAAGTTGGAGACCACCTG	1046		
QY	1771	TTGGAGTCTTGTGGAAAGCTGCAACTTCTAGATAAAATGATGTTCAAACTGAAAGAGCAA	1830		

Db 1047 ACTGAGGCTAGTGGGAAGCTTCACTGCTGGATAAGCTACTAGCATTCCTGTATTCTGGG 1106  
Qy 1831 GGACACAGAGTCTTAATATACACACAGTTTCAGCATATGCTGGACTTACTTTGAAGACTAC 1890  
Db 1107 GGCCATCGGGTTTACTTTTCTCCCAAATGACCCAGATGTTGGATATTTCTCAAGACTAT 1166  
Qy 1891 TGTACCCATAAGAAATGGCAGTACGAGCGAATGTATGGAAGGTTGGCGGAGCTGAGCGG 1950  
Db 1167 ATGGATTACAGAGGCTACAGCTATGAGCGTGTGGATGGTTCTGTGAGAGGAGAAGAGAGA 1226  
Qy 1951 CAAATACGCATAGATCGGTTCAATGCCAAAAATCTTAACAAGTTTGTGTTTGTGCTCTCC 2010  
Db 1227 CACTTGGCCATTAA-----GAACTTTGGACAGCAGCCCATTTTTCGTTTCTCCTGAGT 1280  
Qy 2011 ACAAGAGCTGGTGGCTTAGGAATAAATCTTGCAACGGCTGATACAGTAATCATTTATGAC 2070  
Db 1281 ACTAGGGCAGGTGGAGTTGGCATGAACCTTAACAGCAGCAGATACCTGTGATTTTGTGAC 1340  
Qy 2071 AGTACTGGATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATCGACTTGGCCAA 2130  
Db 1341 AGTACTTTAATCCTCAGAAATGACTTGCAGCAGCTGCCAGGGCTCATCGCATTTGGCCAA 1400  
Qy 2131 ACAAATAAGTGATGATTTATAGGCTCAFAAACCGAGGCACCATTTGAAGAAA 2182  
Db 1401 AACAAAGTCTGTTAAAGTTATTTCGGCTGATTGGTCGAGACACTGTGGAAGAAA 1452

RESULT 8

US-09-671-325-1802  
; Sequence 1802, Application US/09671325  
; Patent No. 6667154  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Bangur, Chaitanya S.  
; APPLICANT: Lodes, Michael A.  
; APPLICANT: Fanger, Gary  
; APPLICANT: Vedvick, Tom  
; APPLICANT: Carter, Darrick  
; APPLICANT: Retter, Marc  
; APPLICANT: Mannion, Jane  
; APPLICANT: Fan, Liqun  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.478C12  
; CURRENT APPLICATION NUMBER: US/09/671,325  
; CURRENT FILING DATE: 2000-09-26  
; NUMBER OF SEQ ID NOS: 1825  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1802  
; LENGTH: 2512  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-671-325-1802

Query Match 5.1%; Score 213.6; DB 4; Length 2512;  
Best Local Similarity 53.8%; Pred. No. 1.9e-51;  
Matches 544; Conservative 0; Mismatches 444; Indels 24; Gaps 4;  
Qy 1180 TTTGATGTCCTCCTCACATCGTATGAGATGATCAACCTAGATTTCAGAGTTCATAAACCA 1239  
Db 456 TTTCAATGTGCTACTGACTACCTATGAGATTGCTTGAAGATGCTATCTATAAATCA 515  
Qy 1240 ATTAAGTGGAGTGCATGATTGTTGATGAAGTGCATCGACTGAAAAATAAGGATTCAAAG 1299  
Db 516 TTCCCTTGGAGTGTCTTGTGTTGGATGAAGCTCACAGGTTGAAAAACCAAGCTCCCTG 575  
Qy 1300 CTGTTCTCTTCAATGACACAGTATTCAAAGTAACCCGCTATTTCTTGACAGGAACCA 1359  
Db 576 CTGCATAAGACCTTGTCAAGATTCTCAGTAGTCTTCAGTCTCCTGTGACCGGAACCTCCC 635  
Qy 1360 CTTCAGAACAACCTGGATGAACCTTTTTCATGCTCATGCACTTTTCTTGATCGGGGAAGTTT 1419

Db 636 ATCAGAAACAGCCTCCAAGAGCTCTACTCCCTCCTCAGTTTGTGGAGCCTGATCTCTTT 695  
Qy 1420 -----GGAAGTTTGGAGGAGTTCCAGGAGGAGTTCAAAGATATTAAATCAAGAGGAGCAG 1473  
Db 696 TCCAAGGAAGAGGTGGAGATTATTATCAACGCTACCAGGATATTGAGAAAGAAATCTGAG 755  
Qy 1474 ATCTCAAG---GTTGCACAAAATGTTGGCTCCACATTTTGCCTCAGAAAGGTAAAAAAGAC 1530  
Db 756 TCAGCAAGTGAACCTGCACAAACTCTTGAGCCATTTCTGCTGAGCGGAGTGAAGCTGAG 815  
Qy 1531 GTAATGAAAGACATGCCCCCCCCAAAAGGAGCTCATTTTGGCTGTGTGATCTGAGCAGTCTG 1590  
Db 816 GTAGCTACAGAGCTTCCCAAGAAGACAGAAAGTAGTGATATACCATGGCATGTCAAGATTG 875  
Qy 1591 CAGAAAGAAATATTACAAAGCTATTTTACCCTGTAATTAACAAGTATTGACAAAAAGGGA 1650  
Db 876 CAGAAAGAAATACTACAAAGGCCATTTTGATGAAAGACCTAGATGCATTTGAAATGAGACG 935  
Qy 1651 GGTGCTCAAAATTTCCCTTAATAAACATTTATGATGGAATTAAGAAAAAGTATGCTGCCATCCT 1710  
Db 936 GCAAAGAAGGTTAAACTACAGAACATTTTGTCCCAGCTTCGAAAGTGTGTGATCACCCCA 995  
Qy 1711 TATATGCTAGAGGTTGTTGAGCCAGTTATTTCAGCAGCGCAAAATGAAGCTTTCAAACAACTT 1770  
Db 996 TATTGTTTGATGGTGTGGAGCC-----GGAGCCTTTTGAAGTTGGAGACCACCTG 1046  
Qy 1771 TTGGAGTCTTGTGGAAAGCTGCAACTTCTAGATAAAATGATGGTCAAACTGAAAGAGCAA 1830  
Db 1047 ACTGAGGCTAGTGGAAAGCTTCACTGCTGGATAAGCTACTAGCATTCCTGTATTCTGGG 1106  
Qy 1831 GGACACAGAGTCTTAATATACACACAGTTTTCAGCATATGCTGGACTTACTTTGAAGACTAC 1890  
Db 1107 GGCCATCGGGTTTACTTTTCTCCCAAATGA CCCAGATGTGGATATTCTCCAAGACTAT 1166  
Qy 1891 TGTACCCATAAGAAATGGCAGTACGAGCGAATTTGATGGAAGGTTGGCGGAGCTGAGCGG 1950  
Db 1167 ATGGATTACAGAGGCTACAGCTATGAGCGTGTGGATGTTCTGTGAGAGGAGAAGAGAGA 1226  
Qy 1951 CAAATACGCATAGATCGGTTCAATGCCAAAAATTTCTAACAAAGTTTGTGTTTGTCTCTCC 2010  
Db 1227 CACTTGGCCATTAA-----GAACTTTGGACAGCAGCCCATTTTCGTTTCTCCTGAGT 1280  
Qy 2011 ACAAGAGCTGGTGGCTTAGGAATAAATCTTGCAACGGCTGATACAGTAATCATTTATGAC 2070  
Db 1281 ACTAGGCAGGTGGAGTTGGCATGAACCTTAACAGCAGCAGATACTGTGATTTTGTGAC 1340  
Qy 2071 AGTACTGGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATCGACTTGGCCAA 2130  
Db 1341 AGTACTTTAATCCTCAGAAATGACTTGCAGCAGCTGCAAGCAGCTGCCAGGGCTCATCGCAATTTGGCCAA 1400  
Qy 2131 ACAAATAAGGTGATGATTTATAGGCTCATAAACCGAGGCACCATTTGAAGAAA 2182  
Db 1401 AACAAAGTCTGTTAAAGTTATTTCGGCTGATTGGTCGAGACACTGTGGAAGAAA 1452

RESULT 9

US-09-579-181-11  
; Sequence 11, Application US/09579181  
; Patent No. 6365372  
; GENERAL INFORMATION:  
; APPLICANT: Chrivia, John  
; APPLICANT: Yaciuk, Peter  
; TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)  
; FILE REFERENCE: 16153-4247  
; CURRENT APPLICATION NUMBER: US/09/579,181  
; CURRENT FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: 60/136,620  
; PRIOR FILING DATE: 1999-05-27  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 11  
; LENGTH: 8916  
; TYPE: DNA

```

; ORGANISM: Human
US-09-579-181-11

Query Match          2.7%; Score 112.2; DB 4; Length 8916;
Best Local Similarity 53.4%; Pred. No. 1.4e-21;
Matches 259; Conservative 0; Mismatches 223; Indels 3; Gaps 1;

QY 1759 TTCAAACAACCTTTGGAGTCTTGTGGAAAGCTGCAAACTTCTAGATAAAATGATGGTCAAA 1818
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5326 TTAAGACTCATCCAGTATGATTGCGAAAGTTGCAGACGTTGGCAGTGTCTGTCGGCAG 5385

QY 1819 CTGAAAGAGCAAGACACAGAGTCTTAATATACACACAGTTTCAGCATATGCTGGACTTA 1878
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5386 CTCAAGGCAGAGGGCCACCGAGTGTCTATCTTCAACCAGATGACCCGAATGCTGGATGTA 5445

QY 1879 CTTGAAGACTACTGTACCCATAAGAAATGGCAGTACGAGCGAATTGATGGAAGGTTGGC 1938
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5446 TTGGAGCAGTTTCTACCTACCATGGCCATCTCTACCTGGCCTGGATGGATCTACTAGA 5505

QY 1939 GGAGCTGAGCGGCAAAATACGCATAGATCGGTTCAATGCCAAAATTTCTAACAGTTTGT 1998
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5506 GTTGAACAGAGACAGGCCTTGATGGAACGGTTCAATGCAGACAA---ACGCATATTCTGC 5562

QY 1999 TTTTGTGCTCTCCACAAGAGCTGGTGGCTTAGGAATAAAATCTTGCAACGGCTGATACAGTA 2058
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5563 TTCATCCTTTCAACTCGAGTGGGGTGTGGCGTGAACTGACAGGACGACACTGTT 5622

QY 2059 ATCATTTATGACAGTGAATCCTCATGTGATCTTCAAGCAATGGCTAGAGCTCAT 2118
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5623 GTTTTTATGACAGGACTGGAATCCCACCATGGATGCTCAGGCCCAGGACCGCTGTAC 5682

QY 2119 CGACTTGGCCAAACAATAAAGTGATGATTTATAGGCTCATAAACCGAGGCACCATGAA 2178
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5683 CGAATTGGCCAGACCCGGGATGCCACATATATAGGCTTATCAGTGAACGGACAGTGGAG 5742

QY 2179 GAAAGGATGATGCAATTGACTAAAAGAAAATGGTTCTAGAGCATCTTGTGTTGGGAAA 2238
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5743 GAGAACATCTTAAAAAAGGCAATCAGAAGAGAAATGTTGGGGACATGGCCATTGAGGGA 5802

QY 2239 CTCAA 2243
    |||
Db 5803 GGCAA 5807

RESULT 10
US-09-579-181-10
; Sequence 10, Application US/09579181
; Patent No. 6365372
; GENERAL INFORMATION:
; APPLICANT: Chrivia, John
; APPLICANT: Yaciuk, Peter
; TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)
; FILE REFERENCE: 16153-4247
; CURRENT APPLICATION NUMBER: US/09/579,181
; CURRENT FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/136,620
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 9354
; TYPE: DNA
; ORGANISM: Human
US-09-579-181-10

```

[illegible]



```
; LENGTH: 1868 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HEARFET01
; CLONE: 1645339
US-09-023-655-429

Query Match          2.5%; Score 103.8; DB 4; Length 1868;
Best Local Similarity 65.1%; Pred. No. 1.4e-19;
Matches 153; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY      1988 ACAAGTTTGTGTTTTTGCTCTCCACAAGAGCTGGTGGCTTAGGAATAAAATCTTCGAACGG 2047
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       55 ATATCTTTGTGTTCTGCTATCAACAAAAGCTGGTGGATTAGGAATAAATCTGACTTCAG 114

QY      2048 CTGATACAGTAATCATTTATGACAGTGACTGGAAATCCTCATGCTGATCTTCAAGCAATGG 2107
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       115 CAAATGTTGTATACTTCACGATATTGACTGTAAATCCTTATAATGACAAACAGCAGAAG 174

QY      2108 CTAGAGCTCATCGACTTGGCCAAAACAANAATAAGTGATGATTTATAGGCTCATAAACCGAG 2167
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       175 ATAGATGCCATAGAGTAGGCCAGACTAAAGAAGTACTAGTTATAAAACTATAAAGCCAAG 234

QY      2168 GCACCATTGAAGAAAGGATGATGCAATTGACTAAAAAGAAAAATGGTTCTAGAGCA 2222
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       235 GGACGATTGAAGATCCATGCTAAAAATTAACCAACAGAAAATTGAAACTAGAAC 289

RESULT 12
US-09-328-111-553/c
; Sequence 553, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 553
; LENGTH: 575
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(575)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-553

Query Match          2.5%; Score 103.2; DB 3; Length 575;
Best Local Similarity 55.2%; Pred. No. 9.5e-20;
Matches 223; Conservative 0; Mismatches 178; Indels 3; Gaps 1;

QY      1241 TTAAGTGGGAGTGCATGATTGTTGATGAAGGTCATCGACTGAAAAATAAGGATTCAAAGC 1300
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       419 TAAATTGGGCATTTATAGTGTTGTTGATGAAGCACACCGGATTAAGAATATGATGACTCCCTTC 360
```

Db 3462 CTGGGGGCTCGGCTGAACCTCCAGTCGGCAGACACTGTGATCATTTTGACAGCGACT 3521  
QY 2078 GGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATCGACTTGGCCAAACAAATA 2137  
Db 3522 GGAATCCTCACCAGGACCTGCAAGCGCAGGACCGAGCCACCGCATCGGCAGCAGAAAG 3581  
QY 2138 AGGTGATGATTTATAGGCTCATAAACCGAGGACCATTTGAAGAAAGGAT 2186  
Db 3582 AGGTGCGTGTGCTCCGCCTCTGCACCGTCAACAGCGTGGAGGAGAAGAT 3630

RESULT 14  
US-09-535-008-66  
; Sequence 66, Application US/095335008  
; Patent No. 6465629  
; GENERAL INFORMATION:  
; APPLICANT: Wong, Alexander K.C.  
; APPLICANT: Tavtigian, Sean V.  
; APPLICANT: Teng, David H.-F.  
; TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE  
; TITLE OF INVENTION: AND OTHER CANCER TYPES  
; FILE REFERENCE: 2318-259  
; CURRENT APPLICATION NUMBER: US/09/535,008  
; CURRENT FILING DATE: 2000-03-23  
; EARLIER APPLICATION NUMBER: U.S. 60/125,806  
; EARLIER FILING DATE: 1999-03-23  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 66  
; LENGTH: 5468  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (75)..(5012)  
; FEATURE:  
; NAME/KEY: conflict  
; LOCATION: (1780)  
; OTHER INFORMATION: GenBank Accession No. 6465629 U29175 shows a C at this  
; OTHER INFORMATION: position (position 1784 in GenBank) rather than  
; OTHER INFORMATION: the G shown here.  
; FEATURE:  
; NAME/KEY: allele  
; LOCATION: (26)  
; OTHER INFORMATION: Polymorphism of either T or C in this noncoding  
; OTHER INFORMATION: region.  
; FEATURE:  
; NAME/KEY: allele  
; LOCATION: (1583)  
; OTHER INFORMATION: Polymorphism of A or G resulting in a silent  
; OTHER INFORMATION: mutation.  
; FEATURE:  
; NAME/KEY: allele  
; LOCATION: (1598)  
; OTHER INFORMATION: Polymorphism of T or C resulting in a silent  
; OTHER INFORMATION: mutation.  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (4498)..(4499)  
; OTHER INFORMATION: A CAG is missing between these bases as compared  
; OTHER INFORMATION: to SEQ ID NO:1.  
US-09-535-008-66

Query Match 2.4%; Score 100.2; DB 4; Length 5468;  
Best Local Similarity 52.8%; Pred. No. 3.2e-18;  
Matches 216; Conservative 0; Mismatches 193; Indels 0; Gaps 0;

QY 1778 CTTGTGAAAGCTGCAACTTCTAGATAAAATGATGGTCAAACTGAAAGAGCAAGACACA 1837  
Db 3307 CCTCGGTAAATTTGAGCTTCTTGATAGAAATTTCCCAAACCTCCGAGCAACCAACCACA 3366  
QY 1838 GAGTCCTAATATACACACAGTTTCAGCATATGCTGGACTTACTTGAAGACTACTGTATCCC 1897  
Db 3367 AAGTGTCTGTCTTGCCAAATGACCTCCCTCATGACCATGGAAGATTACTTTGCGT 3426  
QY 1898 ATAAGAAATGGCAGTACGAGCGAATTGATGGAAGGTTGGCGGAGCTGAGCGGCAAAATAC 1957  
Db 3427 ATCGCGGCTTTAAATACCTCAGGCTTGATGGAACCAACGAGGCGGAGGACCGGGGCATGC 3486  
QY 1958 GCATAGATCGGTTCAATGCCAAAAAATTCTAACAAAGTTTGTGCTCTCCACAAGAG 2017  
Db 3487 TGCTGAAAAACCTTCAACGAGCCCCGGCTCTGAGTACTTTCATCTTCTGCTCAGCACCCGGG 3546  
QY 2018 CTGTCGGCTTAGGAATAAATCTTCAACGGCTGATACAGTAATCATTTATGACAGTGACT 2077  
Db 3547 CTGGGGGCTCGGCCTGAACCTCCAGTCGGCAGACACTGTGATCATTTTGTGACAGCGACT 3606  
QY 2078 GGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATCGACTTGGCCAAACAAATA 2137  
Db 3607 GGAATCCTCACCAGGACCTGCAAGCGCAGGACCGAGCCACCGCATCGGCGCAGAGAAG 3666  
QY 2138 AGGTGATGATTTATAGGCTCAFAAAACCGAGGCACCATTTGAAGAAAGGAT 2186  
Db 3667 AGGTGCGTGTGCTCCGCCTCTGCACCGTCAACAGCGTGGAGGAGAAGAT 3715

RESULT 15  
US-09-535-008-1  
; Sequence 1, Application US/095335008  
; Patent No. 6465629  
; GENERAL INFORMATION:  
; APPLICANT: Wong, Alexander K.C.  
; APPLICANT: Tavtigian, Sean V.  
; APPLICANT: Teng, David H.-F.  
; TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE  
; TITLE OF INVENTION: AND OTHER CANCER TYPES  
; FILE REFERENCE: 2318-259  
; CURRENT APPLICATION NUMBER: US/09/535,008  
; CURRENT FILING DATE: 2000-03-23  
; EARLIER APPLICATION NUMBER: U.S. 60/125,806  
; EARLIER FILING DATE: 1999-03-23  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 5471  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (75)..(5015)  
; FEATURE:  
; NAME/KEY: conflict  
; LOCATION: (1780)  
; OTHER INFORMATION: GenBank Accession No. 6465629 U29175 shows a C at this  
; OTHER INFORMATION: position (position 1784 in GenBank) rather than  
; OTHER INFORMATION: the G shown here.  
; FEATURE:  
; NAME/KEY: allele  
; LOCATION: (26)  
; OTHER INFORMATION: Polymorphism of either T or C in this noncoding  
; OTHER INFORMATION: region.  
; FEATURE:  
; NAME/KEY: allele  
; LOCATION: (1583)  
; OTHER INFORMATION: Polymorphism of A or G resulting in a silent  
; OTHER INFORMATION: mutation.  
; FEATURE:  
; NAME/KEY: allele  
; LOCATION: (1598)  
; OTHER INFORMATION: Polymorphism of T or C resulting in a silent

; OTHER INFORMATION: mutation.  
; FEATURE:  
; NAME/KEY: allele  
; LOCATION: (1892)  
; OTHER INFORMATION: Polymorphism of A or G resulting in a silent  
; OTHER INFORMATION: mutation.  
US-09-535-008-1

Query Match		2.4%	Score 100.2;	DB 4;	Length 5471;
Best Local Similarity		52.8%;	Pred. No. 3.2e-18;		
Matches	216;	Conservative	0;	Mismatches	193;
				Indels	0;
				Gaps	0;
Qy	1778	CTGTGTGAAAGCTGCAACTTCTAGATAAAATGATGGTCAAACTGAAAGAGCAAGGACACA	1837		
Db					
Qy	3307	CCTCGGGTAAATTGAGCTTCTTGATAGAAATCTTCCAAACTCCGAGCAACCAACCACA	3366		
Db					
Qy	1838	GAGTCCTAATATACACACAGATTTTCAGCATATGCTGGACTTACTTTGAAGACTACTGTACCC	1897		
Db					
Qy	3367	AAGTGTGCTGTTCTGCCAAATGACCTCCCTCATGACCATCATGGAAGATTACTTTGCGT	3426		
Db					
Qy	1898	ATAAGAAATGGCAGTACGAGCGGAATTGATGAAAGGTTGGCGGAGCTGAGCGGCAATAC	1957		
Db					
Qy	3427	ATCGCGGCTTTAAATACCTCAGGCTTGATGGAACCAACGAAGCGGGAGACCGGGGCATGC	3486		
Db					
Qy	1958	GCATAGATCGGTTCAATGCCAAAATTCTAACAAGTTTGTGTTTTTTGCTCTCCACAAGAG	2017		
Db					
Qy	3487	TGCTGAAAACCTTCAACGAGCCCGGCTCTGAGTACTTTCATCTTCCCTGCTCAGCACCCGGG	3546		
Db					
Qy	2018	CTGGTGGCTTAGGAATAAATCTTGCAACGGCTGATACAGTAATCATTTATGACAGTGACT	2077		
Db					
Qy	3547	CTGGGGGCTCGGCCTGAAACCTCCAGTCGGCAGACACTGTGATCAITTTTGACAGCGACT	3606		
Db					
Qy	2078	GGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATCGACTTGGCCAAAACAAATA	2137		
Db					
Qy	3607	GGAATCCTCACCCAGGACCTGCAAGCGCAGGACCGAGCCACCCGCATCGGGCAGCAGAACG	3666		
Db					
Qy	2138	AGGTGATGATTTATAGGCTCATAAACCGAGGCACCATTGAAGAAAGGAT	2186		
Db					
Qy	3667	AGGTGCGTGTGCTCCGCCCTCTGCACCCGTCAACAGCGTGGAGGAGAAGAT	3715		
Db					

Search completed: September 19, 2004, 02:55:47  
Job time : 274 secs



**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 20:33:51 ; Search time 1790 Seconds  
(without alignments)  
11778.315 Million cell updates/sec

Title: US-10-049-137-1  
Perfect score: 4177  
Sequence: 1 atgagtagtttggtggagag.....caactggtaaatcaagattc 4177

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3327077 seqs, 2523723180 residues

Total number of hits satisfying chosen parameters: 6654154

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq2:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1145.2	27.4	4551	17	US-10-675-072A-42 Sequence 42, Appl
2	1038.8	24.9	4459	17	US-10-437-963-53427 Sequence 53427, A
3	850.8	20.4	3204	9	US-09-938-842A-2526 Sequence 2526, Ap
4	850.8	20.4	3204	11	US-09-938-842A-2526 Sequence 2526, Ap
5	434.2	10.4	2578	13	US-10-424-599-91500 Sequence 91500, A
6	421.4	10.1	5739	9	US-09-960-253-142 Sequence 142, App
7	421.4	10.1	6417	9	US-09-962-436-288 Sequence 288, App
8	421.4	10.1	6417	15	US-10-096-534-28 Sequence 28, Appl
9	421.4	10.1	6475	15	US-10-037-270-325 Sequence 325, App
10	421.4	10.1	6475	16	US-10-117-722-325 Sequence 325, App
11	420.8	10.1	6331	15	US-10-175-523-51 Sequence 51, Appl
12	420.8	10.1	6331	16	US-10-159-563-176 Sequence 176, App
13	420.6	10.1	6337	16	US-10-062-674-1723 Sequence 1723, Ap
14	409.6	9.8	7805	14	US-10-044-090-370 Sequence 370, App

15	409.6	9.8	7805	15	US-10-006-285-400 Sequence 400, App
16	400.4	9.6	654	13	US-10-424-599-23261 Sequence 23261, A
17	387	9.3	3716	16	US-10-104-047-476 Sequence 476, App
18	387	9.3	4675	17	US-10-115-635-216 Sequence 216, App
19	387	9.3	9646	13	US-10-462-261-1 Sequence 1, Appli
20	321.2	7.7	7001	17	US-10-437-963-98673 Sequence 98673, A
21	316.2	7.6	612	13	US-10-424-599-111448 Sequence 111448, Sequence 60603, A
22	315.4	7.6	620	13	US-10-424-599-60603 Sequence 10, Appl
23	296.6	7.1	6608	8	US-08-973-363-10 Sequence 140, App
24	285.4	6.8	6364	14	US-10-071-766-140 Sequence 21548, A
25	285.4	6.8	7764	10	US-09-814-353-21548 Sequence 60, Appl
26	285.4	6.8	7764	16	US-10-341-434-60 Sequence 13362, A
27	272.2	6.5	1862	13	US-10-425-114-13362 Sequence 1579, Ap
28	272.2	6.5	3262	13	US-10-412-699B-1579 Sequence 84139, A
29	269	6.4	1979	13	US-10-424-599-84139 Sequence 25462, A
30	266.2	6.4	4407	16	US-10-369-493-25462 Sequence 1, Appli
31	264	6.3	1874	14	US-10-005-057A-1 Sequence 1, Appli
32	264	6.3	1874	17	US-10-675-072A-1 Sequence 1, Appli
33	259.6	6.2	3120	13	US-10-058-024-5 Sequence 5, Appli
34	259	6.2	3594	17	US-10-437-963-51978 Sequence 51978, A
35	253.4	6.1	3743	17	US-10-437-963-49246 Sequence 49246, A
36	252.6	6.0	3192	9	US-09-938-842A-2271 Sequence 2271, Ap
37	252.6	6.0	3192	11	US-09-938-842A-2271 Sequence 2271, Ap
38	249.8	6.0	2195	17	US-10-343-710-115 Sequence 115, App
39	249.8	6.0	6222	16	US-10-120-988-114 Sequence 114, App
40	248.2	5.9	3548	13	US-10-425-114-31425 Sequence 31425, A
41	248.2	5.9	4011	13	US-10-412-699B-1583 Sequence 1583, Ap
42	247	5.9	4221	16	US-10-369-493-25997 Sequence 25997, A
43	245.4	5.9	3174	10	US-09-934-455-431 Sequence 431, App
44	245.4	5.9	3174	13	US-10-412-699B-789 Sequence 789, App
45	245.4	5.9	3174	13	US-10-412-699B-1845 Sequence 1845, Ap

ALIGNMENTS

RESULT 1

US-10-675-072A-42  
; Sequence 42, Application US/10675072A  
; Publication No. US20040098760A1  
; GENERAL INFORMATION:  
; APPLICANT: Yumin, Tao  
; APPLICANT: Gordon-Kamm, William  
; APPLICANT: Shen, Bo  
; APPLICANT: Lowe, Keith  
; APPLICANT: Danilevskaya, Olga  
; APPLICANT: Mahajan, Pramod  
; APPLICANT: Rafalski, Jan Antoni  
; APPLICANT: Sakai, Hajime  
; APPLICANT: Klein, Theodore  
; TITLE OF INVENTION: Transcriptional Regulatory Nucleic Acids, Polypeptides, and Methods of Use Thereof  
; FILE REFERENCE: 1288R  
; CURRENT APPLICATION NUMBER: US/10/675,072A  
; CURRENT FILING DATE: 2003-09-30  
; PRIOR APPLICATION NUMBER: 10/005,057  
; PRIOR FILING DATE: 2001-12-04  
; PRIOR APPLICATION NUMBER: 60/251,555  
; PRIOR FILING DATE: 2000-12-06  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 42  
; LENGTH: 4551  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (343)...(4332)  
US-10-675-072A-42

Query Match 27.4%; Score 1145.2; DB 17; Length 4551;  
Best Local Similarity 62.1%; Pred. No. 3e-295;  
Matches 2098; Conservative 0; Mismatches 1193; Indels 87; Gaps 15;





Db 2485 TTAGGACAGACTAGTAAGGTGATGATATACAGGCTTGTAGCCGAGGTACAATTGAGGAG 2544  
QY 2182 AGGATGATGCAATTGACTAAAAAGAAAATGTTTCTAGAGCATCTTGTGTTGGGAAACT- 2240  
Db 2545 CGAATGATGCAGCTTACAAAAAGAAAGATTTATTGGAGCACTTAGTTGTTGGTCGACTC 2604  
QY 2241 --CAAAACACAAAACATTAAATCAGGAAGAGTTAGATGACATCATCAGGTATGGATCAAAG 2298  
Db 2605 ACCAAAGCTAATAATGTCAATCAGGAGGAGTTGGATGATATTATACGCTATGGATCAAAG 2664  
QY 2299 GAGCTTTTGTCTAGTGAAGATGATGAAGCAGGAAAGTCTGGAAAAATTCATTATGATGAT 2358  
Db 2665 GAGCTTTTGAAGACGAGAAATGATGAA-----TCTCGCCAAATTCATTATGACGAA 2715  
QY 2359 GCGGCTATAGACAAAATTGCTTGATCGTGTATCTCGTGGAGGCAGAGGAAGTCTCAGTGGAT 2418  
Db 2716 GCTGCAATTGAGAGGTTGTTAGACCGTGTATCAAGTTGACCGGTGATGAA---TCTGTGGAA 2772  
QY 2419 GATGAAGAGGAGAAATGGATTCTTTAAAGGCTTTCAAGGTGGTAAATTTGAATATATAGAT 2478  
Db 2773 GATGAAGAAGAGATGGATTCTTTAAAGGATTCAAAGGTTGCAAACTTTGAATATATCGAT 2832  
QY 2479 GAAAATGAGGCAGCAGCATTAGAGGCACAGAGATCGTGTCTGTAAGCAAAATCTTCAGCA 2538  
Db 2833 GAGGCAAAGSCTCAGGCAGAAAAAGAGGAG-----GCACGGAGAAAGGCTGCAGCTGAG 2886  
QY 2539 GGCAATTCTGATAGACCAAGTTATTGGGAAGAGTTGTTAAAGATATAAATTTGAGCTGCAC 2598  
Db 2887 GCTGAAAAATTCGAAAGAAACTACTGGGATGAACTATTGAAGGATAGATATGATGTACAG 2946  
QY 2599 CAGGCTGAGGAGCTTAATGCTCTTGGAAAAAGGAAGAGTCGCAAGCAGTGGTATCC 2658  
Db 2947 AAAGTTGAAGAAACATACTGCTATGGGAAAAAGGAAAAAGCCGCAACAGATGGCTGCC 3006  
QY 2659 ATTGAAGAAGATGATCTTGTGTTTGGAAAGATGTGAGCTCTGATGGAGATGAAAGTTAT 2718  
Db 3007 GCTGATGAAGATGACATTCAATGATTTAAGTTCCGAA-----GATGAGGATTACTCATTTG 3060  
QY 2719 GAAGCTGAGTCAACAGATGGTGAAGCAGCAGGACAAGGAGTTTCAGACGGGTTCGACGGCCG 2778  
Db 3061 GAGGATGACATTTTCAGATAATGACACAAAGTTGCAAGGAAATATTTCTGGGAAGAGGGGA 3120  
QY 2779 TACAGAAGAAAGSGTCGCGA--TAATTTGGAACCAACTCCGTTGATGGAAGGTGAGGGG 2835  
Db 3121 CAATATTCTAAGAGAAAAATCACGTAATGTGATTCTATTCCATTGATGGAGGCGGAAGGA 3180  
QY 2836 AGATCTTTCAGAGTACTGGGTTTCAACCAAGTCAAAAGGGCCATTTTGTACAGACTTTG 2895  
Db 3181 CGTACCTTGAGAGTCTTGGATTCAACCATGCTCAACGAGCAATGTTCTACAGACACTC 3240  
QY 2896 ATGAGGTAATGGAGTGGCAATTTTGATGGAGGAGTTTGTTCCTCGCTTAAAGCAGAAG 2955  
Db 3241 AATAGATTCCGTTTTCAGAAATTATGACTGGAAAGAGTATCTTCCTCGTCTTAAAGGAAAA 3300  
QY 2956 ACCTTTGAAGAAATAAATGAATATGGAATACTCTTTGAAGCACATTTGCTGAAGAAATA 3015  
Db 3301 AGTGTGAGGAAATCCAGAGATATGCTGAACTTGTCAATGGCACATCTTGTGAAGAAAT 3360  
QY 3016 GACGAGAAATTCCTCAACCTTTTCAGATGGTGTGCCAAAGGAAGGACTTAGAATAGAAGAT 3075  
Db 3361 AATG--ATTCTGACTATTTTTCAGATGGCGTTCCAAAGGAAATGATGCGTGTGATGAT 3417  
QY 3076 GTTCTAGTCAGAAATTGCTCTTCTGATACCTAGTTCAGGAGAAAGGTGAATTTGTAGAAGAT 3135  
Db 3418 GTACTAGTCAGGATAGCAAAACATATCCCTTATCGAGGAGAAAGATGGTCCACAGGACCA 3477  
QY 3136 CATCCAGGGAAACCTGTTTCCCTCTCGCAATCTTTGAAAGATTCCTCGGACTGAGAAAT 3195  
Db 3478 GGAAGAAATTACAAACATTTTCTCTAATTACTTGCTCTATGAGTTCCAAGGCTTATCTGGT 3537  
QY 3196 GGAAGAAATTGGAAGGAGGAACATGACAAGATAATGATACGTGCTGTTTAAAGCATGGG 3255

Db 3538 GGAAGAAATATGAAAGCGGAGCATGATCTACTGTTACTGAGAGGCATACTGAAGCATGGA 3597  
QY 3256 TACGACGGTGGCAAGCTAATTGTTGATGACAAAGAGTTGGGGATCCAAGAGCTTATCTGC 3315  
Db 3598 TATGCAAGGTGGCAGTATATATCAGATGACAGAGAGAAATGGGCTTTTGTAGGCTGCACGA 3657  
QY 3316 AAAGAAATTGAATTTCCCT 3333  
Db 3658 CGAGAGCTGCATCTCCCT 3675  
RESULT 2  
US-10-437-963-53427  
; Sequence 53427, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 53427  
; LENGTH: 4459  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_55629C.1  
US-10-437-963-53427  
Query Match 24.9%; Score 1038.8; DB 17; Length 4459;  
Best Local Similarity 61.2%; Pred. No. 9.6e-267;  
Matches 1978; Conservative 0; Mismatches 1067; Indels 189; Gaps 11;  
QY 280 GTTAGTCTCTTTAACGAGATAGATAAGATATTGGATTGT---GAAATGCGTCTCTACAAAA 336  
Db 160 GTTAGCCCTCTTACGGAAATGAAAAAGATATTAGACTGTGAAGAAACAAAACCTGATGCT 219  
QY 337 TCTAGTGAACAAGGTTCTCCGATCGGAACCGAACCGAAGCCAAATTTTGTGAAACAGTATCTC 396  
Db 220 TCTGAAGAGACTAGTTCTTCTGAGTCCGGATCAAAGAAAGAAACCTGTCAAACGATATCTT 279  
QY 397 GTGAAGTGAAGGGATTATCATACCTTCACTGCTCTTGGTGCTTGAGAAGGAGTTCCAG 456  
Db 280 ATAAATGGAAGGAATATCACCTTCACTGCACCTTGGGTTTCAGAAAGTGAATATTTG 339  
QY 457 AAGGCTTATAAGTCAAATCATCGTTTAAAAACCAGAGTGAACAAATTTTCACCGTCAAATG 516  
Db 340 GAACTGCCAAGATATACCCCGCTGAAAACCTAGACTCAATAACTTCCATAAGCAAATG 399  
QY 517 GAGTCATTCAAATAACAGCGAAGATGATTTTGTGGCATACGTCTCCTGAGTGGACCCTGTT 576  
Db 400 GATTCAACGGATAAGTCTGATGATGACTATTCTGCAATTAGACCTGAGTGGACTACTGTT 459  
QY 577 GATCGGATTTGCTTCGCTGCAGAGA---GGAAGATGGGAGCTGGAATATCTTGTCAAATAT 633  
Db 460 GACAGGATCCTTGCTACCAGAAAAAGCTCTACTGGTGAAGGAGTACTATGTGAAATGG 519  
QY 634 AAAGAGCTATCCTATGATGAATGTTATTGGGAGTCAGAAATCAGACATCTCAACCTTCCAG 693  
Db 520 AAGGAACTAACATACGACGAATGACTTGGGAAAATGACTCCGACATTCGAGTCTTTTCAG 579  
QY 694 AATGAAATTCAAAGGTTCAAGGATGTAAATTTCTAGAACTCCAGAAAGTAAAGATGTTGAC 753  
Db 580 CCTCAGATTGAACGCTTCAATGAGATTCTAGTCCAGGAGGAAGAGTCTACTGACAAGTGC 639



```

Db      2797 AGTTTACAATCAGGCCTTGCTGGGAGGAGGGGCCCATATTCCAAGAAAAACAACGTAGT 2856
Qy      2800 AATTTGGAACCAACTCCGTTGATGGAAGGTGAGGGGAGATCTTTCAGAGTACTGGGTTTC 2859
Db      2857 AATGTTGATTCACTTCCATTCAATGAGGGCGAAGGACGGGCTTTGAGAGTTTATGGGTTT 2916
Qy      2860 AACCAGAGTCAAGGGCCATTTTGTACAGACTTTTGATGAGGTATGGAGCTGGCAATTTT 2919
Db      2917 AACCAAAATTCAGCGAACACAATTCCCTTCAACCGCTTATGAGATATGGTTTTCAGAACTAC 2976
Qy      2920 GATTGGAAGGAGTTTGTTCCTCGCTTAAAGCAGAAAGACCTTTGAGAAATAAATGAATAT 2979
Db      2977 GATTGGAAGGAATTACTCCTCGATTGAAGGGGAAAAAGTTGTGAGGAAATCCAGAGATAT 3036
Qy      2980 GGAATACTCTTCTTGAAGCACATTGCTGAAGAAATAGACGAGAATTCTCCAACCTTTTCA 3039
Db      3037 GCTGAACCTTGTATGATCCATCTTCTTGAGGACATCAATGAC---TCAGGATATTATGCA 3093
Qy      3040 GATGGTGTGCCAAGGAAGGACTTAGAATAGAAATGTTCTAGTCAGAAATTGCTCTTCTG 3099
Db      3094 GATGGCGTGCCAAAGGAAATGC---GCACAGATGAGACATTTGGTCAGGCTAGCCAAACATA 3150
Qy      3100 ATACTAGTTCAGGAGAAGGTGAAATTTGTAGAAGATCATCCAGGGAACCTGTTTTCCTCC 3159
Db      3151 TCACCTTGTTGAAGAGAAGGTGGCTGCCATGGAACAAGGAAAAATAACAAAACTCTTTCCC 3210
Qy      3160 TCTCGCATCTTGAAAAGATTCCCGGACTGAGAAGTGGAATAATTGGAAAGGAGGAACAT 3219
Db      3211 AGCTATTTGTTATATGAATTTCTTAGCTTAGTTGGTGAAGAGTATGGAAGCGGAACAG 3270
Qy      3220 GACAAGATAATGATACGTGCTGTTTTTAAGCATGGGTACGGACGGTGGCAAGCTATTGTT 3279
Db      3271 GATCTACTGTTGCTGAAAGCATTAATAAAGCATGGGTATGCAAGGTGGCAATATATATCA 3330
Qy      3280 GATGACAAAGAGTTGGGGATCCAAAGAGCTTATCTGCAAAGAATTGAATTTCCCT 3333
Db      3331 GATGACAGAGATAATGGGATTTTGGAGCTGCACGACGACAGAGCTGAGGCTTCCT 3384

RESULT 3
US-09-938-842A-2526
; Sequence 2526, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIP1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2526
; LENGTH: 3204
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-2526

Query Match      20.4%; Score 850.8; DB 9; Length 3204;
Best Local Similarity 61.0%; Pred. No. 1.9e-216;
Matches 1892; Conservative 0; Mismatches 697; Indels 513; Gaps 11;

Qy      280 GTTAGTCTCTTAACGAGATAGATAAGATATGGATTGTGAAATGCGTCCTACAAAATCT 339
```

```

Db      121 GATTGTCTCTTTGGCGAGATTGAAAAAGATTTTGGACCGTGAATGGCGTCTACTGCACT 180
Qy      340 AGTGAACAAAGGTTCTCTCCGAT--GCGGAACCGAAGCCAATTTTGTGAAACAGTATCTC 396
Db      181 AACAAATCCGAATTCTTCTGACAAACGGGACACCAACCCCTAGTTGTCGTGAAACAATATCTT 240
Qy      397 GTGAAGTGAAGGGATTATCATACCTTCACTGCTCTTGGTGCTGAGAAGGAGTTCAG 456
Db      241 GTGAAGTGAAGGGTTATCGTACTTGCACTGCTCTTGGTGCTCCGAGCAGGAGTTCGAG 300
Qy      457 AAGCCTTATAAGTCAAAATCATCGTT-----TAAAAACCCAGAGTGAACAAATTTTCACCGT 510
Db      301 AAGGCTTACAAGTCCCATCCATTTAAAACTAAAAATTAAGGTGACCAGGTTTAAACGCT 360
Qy      511 CAAATGGAGTCATTCA-----ATAACAGCGAAGATGATTTTGTGCTCCATACGTCCTGAG 564
Db      361 GCTATGGACGTATTTCATCGCTGAGAACGGAGCTCATGAGTTATCGCCATTTCGTCCTGAA 420
Qy      565 TGGACCACTGTTGATCGGATTCTTGCTGCAGAGAGGAGAGTGGGAGCTGGAATATCTT 624
Db      421 TGGAAAACTGTTGACCGGATTATTGCTTGCCGGGAAGGAGATGATGGGAGGAGTATCTG 480
Qy      625 GTCAAAATATAAAGAGCTATCCTATGATGAATGTTATTGGSAGTCAGAATCAGACATCTCA 684
Db      481 GTGAAGTACAAAGAACTTTCTTACAGAAATAGTTACTGGAATCTGAATCTGACATCTCA 540
Qy      685 ACCTTCCAGAAATGAAATTTCAAAGGTTCAAGGATGTAATTTCTAGAACTCGCAGAAAGTAAA 744
Db      541 GATTTCAGAAATGAAATCCAAAGGTTTAAGGATATAAACTCTAGCTCTCGTAGAGATAAA 600
Qy      745 GATGTTGACCATATAAAGAAATCCCAGAGACTTTTCAACAGTTTGTATCATACTCCTGAATTC 804
Db      601 TATGTCGAAAAATGAGAGAAATCGAGAGGAATTCAAAACAGTTTGATCTTACTCCTGAATTT 660
Qy      805 CTCAAAGGCTTGTTACATCCATACCAGCTTGAGGGACTTAAATTTTTCGGTTCTCGTGG 864
Db      661 CTTACAGGCACATTGCATACATACCAGCTCGAAGGGCTGAATTTCTTAAGGTATTTCATGG 720
Qy      865 TCAAAAACAGACGCATGTAATCCTTGCTGATGAAATGGGACTAGGCAAGACAATTTCAAAGC 924
Db      721 TCCAAGAAAACCAATGTAATCCTTGCGGATGAAATGGGACTAGGCAAACTATTTCAGAGC 780
Qy      925 ATTGCCCTTTTAGCTTCACITTTTGAGGAGAACCTCATTCGGCATTTGGTAATTGCTCCT 984
Db      781 ATTGCCCTTCTTGGCCTCTTTGTTGAGGAAAACTTATCTCTCATCTAGTTGTTGCTCCT 840
Qy      985 CTATCGACTCTGCGTAACTGGGAGAGAGAGTTTGCACATGCGGCCCCACAGATGAACGTG 1044
Db      841 CTCTCAACCATCCGTAACCTGGGAGAGGGAAATTTGCCACCTGGGCACCCACATATGAATG 900
Qy      1045 GTTATGTATTTTGGCACTGCGCAAGCTCGAGCAGTTTATCAGAGAAACATGAGTTTTACTTA 1104
Db      901 GTTATGTACACTGGGGATTTCGGAAGCACGTGATGTTATATATGGAACAACATGAGTTCTACTTC 960
Qy      1105 TCGAAAAGATCAAAAAAGATCAAGAAAAAGAAATCTGGACAAAATAAGTAGCGAAAGCAAG 1164
Db      961 TCTGAAGGGCGTAAAAGGTTAAAG----- 984
Qy      1165 CAAAAAAGAATCAAGTTTGTATGTCCTCCTCACATCGTATGAGATGATCAACCTAGATTCA 1224
Db      985 ----- 984
Qy      1225 GCAGTTCTAAAAACCAATTAAAGTGGGAGTGCATGATTGTTGATGAAGGTCAATCGACTGAAA 1284
Db      985 ----- 984
Qy      1285 AATAAGGATTCAAAGCTGTTCTCTTTCATTGACACAGTATTAAGTAACCACCGTATTCTT 1344
Db      985 ----- 984
Qy      1345 CTGACAGGAACACCACCTTCAGAACAACTTGGATGAACITTTTCATGCTCATGCAATTTTCTT 1404
Db      985 ----- 984
```



QY	1405	GATCGGGGAAGTTTGGAAAGTTTGGAGGAGTTCAGGAGGAGTTCAAAGATATTAATCAA	1464
Db	985	-----	984
QY	1465	GAGGAGCAGATCTCAAGGTTGCACAAAATGTTGGCTCCACATTTGCTCAGAAAGGTAAAA	1524
Db	985	-----AAA	987
QY	1525	AAAGACGTAATGAAGAAGACATGCCCCCCAAAAGGAGCTCATTTTGCGTGTTGATCTGAGC	1584
Db	988	GACGTGCTCAAGGATAAGGTGCCCCCAAAGAGGAGCTCATTTTGCGTGTTGATATGAGC	1047
QY	1585	AGTCTGCAGAAAGAAATATTACAAAGCTATTTTACCCGTAATTATCAAGTATTGACAAA	1644
Db	1048	AGCCAGCAAAAAGAACTGTACAAAGCTGTTATTACCAATAATTATCAAGTCTTGACAAAG	1107
QY	1645	AAGGAGGTGCTCAAAATTTCCCTTAATAACATATTGATGGAATTACGAAAAGTATGCTGC	1704
Db	1108	AAACGAGATGCTAAAAATATC-----TAACGTGCTAATGAAATTACGCCAAGTTTGCTCT	1161
QY	1705	CATCCTTATATGCTAGAGGGTGTGAGCCAGTTATTACGACGCAAAATGAAGCTTTCAAA	1764
Db	1162	CATCCGTATCTGTTACCAGACTTTGAGCCAAAGATTTGAAGACGCAAAATGAAGCTTTTACA	1221
QY	1765	CAACTTTTGGAGTCTTGTGAAAAGCTGCAACTCTAGATAAAAATGATGGTCAAACTGAAA	1824
Db	1222	AAACTATTGGAAGTTCGCGAAAGTTGCAACTGTAGACAAAATGATGGTCAAACTCAA	1281
QY	1825	GAGCAAGGACACAGAGTCCTAATAATACACACAGTTTCAGCATATGCTGGACTTACTTGAA	1884
Db	1282	GAGCAGGGGCACAGAGTCTAATCTATACCCAAATTCACACACCCCTTACCTTTTGGAA	1341
QY	1885	GACTACTGTACCCATAAGAAATGGCAGTACGAGCAATTTGATGAAAAGTTGGCGGAGCT	1944
Db	1342	GACTACTTCACTTTCAGAAATTGGAATTATGAACGGATTGATGGAAGATCAGTGGACCT	1401
QY	1945	GAGCGGCAAAATACGCATAGATCGGTTCAATGCCAAAAATTCCTAACAGTTTTGTTTTTG	2004
Db	1402	GAGAGGCAAGTAAGAAATAGATCGGTTTAAATGCAGAAAACTCGAACCGGTTCTGTTTTTG	1461
QY	2005	CTCTCCACAAGAGCTGGTGGCTTAGGAATAAATCTTGCAACGGCTGATACAGTAATCATT	2064
Db	1462	CTCAGTACAAGAGCTGGTGGAAATAGGGATAAACCTTGCAACAGCTGATACAGTCATCATT	1521
QY	2065	TATGACAGTGACTGGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATCGACTT	2124
Db	1522	TATGACAGTGATTGGAAATCCTCATGCTGATCTCCAGGCTATGGCTAGAGTTCATCGGCTT	1581
QY	2125	GGCCAAACAAATAAGGTGATGATTTATAGGCTCATAAACCGAGGCACCATTTGAAGAAAGG	2184
Db	1582	GGCCAAACAAATAAGGTAATGATCTACAGACTGATTCACAAAGGTACTGTAGAAGACGG	1641
QY	2185	ATGATGCAATTGACTPAAAAAGAAATGGTTCTAGAGCATCTTGTGTTGGGAAACTCAA	2244
Db	1642	ATGATGGAGATAACCAAGAACAAAGATGCTTCTAGAGCATCTCGTCGTGGG-----A	1692
QY	2245	ACACAAAACATTAATCAGGAAGAGTTAGATGACATCATCAGGTATGGATCAAAGGAGCTT	2304
Db	1693	AAACAACATCTTTGTGAGGACGAGTTTAGATGACATCATCAAATATGGTTCCAAGGAACTT	1752
QY	2305	TTTGCTAGTGAAGATGATGAAGCAGGAAAGTCTGGAATAAATTCATTATGATGATCGGCT	2364
Db	1753	TTCTCTGAAGAGAATGATGAAGCAGGAGGTCTGGAATAAATCCATTACGATGATGCTGCT	1812
QY	2365	ATAGACAAATTGCTTGATCGTGATCTCGTGGAGGCAGAGGAAGTCTCAGTGGATGATGAA	2424
Db	1813	ATCGAACAGTTGCTAGATCGTAATCATGTAGATGCTGTGGAGGTCTCACTGGATGATGAG	1872
QY	2425	GAGGAGAATGGATTCTTAAAGGCTTTCAAGGTGGCTAAATTTTGAATATATAGATGAAAAT	2484
Db	1873	GAGGAAACTGATTTTAAAGAATTTCAAGGTGGCTAGTTTGGAGTATGATGATGAA	1932

QY	2485	GAGGAGCAGCATTAGAGGCACAGAGAGTCGCTGTGAAAGCAAAATCTTCAGCAGGCAAT	2544
Db	1933	AATGAGGCTGCAGCATTAGAGGAGGCACAAGCTATAGAAAACAATCTTCAGTAAGAAAT	1992
QY	2545	TCTGATAGAGCAAGTTATTGGGAAGAGTTGTTAAAAAGATAAAATTTGAGCTGCACCAGGCT	2604
Db	1993	GCTGATAGAACAAGTCACTCGAAAAGATTGCTGAAAAGACAAATATGAAGTGCAACAAGCT	2052
QY	2605	GAGGAGCTTAATGCTCTTGGA AAAAGGAAGAGAAAGTCGCAAGCAGTTGGTATCCATTGAA	2664
Db	2053	GAGGAGCTCAGTCTCTTGGA AAAAGGAAGAGAAACGGCAAGCAGGTGATGTA--TGCT	2109
QY	2665	GAAGATGATCTTGCTGGTTTGGAAAGATGTGAGCTCTGATGGAGATGAAAAGTTATGAAGCT	2724
Db	2110	GAAGATGATCTTGATGGTTTAGAAGAAATCTCTGATGAGGAGACGAATATTGTCTTGAT	2169
QY	2725	GAGTCAACA-----GATGGTGAAGCAGCAGGACAA	2754
Db	2170	GATTTAAAAAGTGACCTCTGATGAGGAGGAGGAAGCAGATGAACCCGAAAGCAGCTCGACAG	2229
QY	2755	GGAGTTC--AGACGGGTGCGACGCCGTACAGAAAGAAAGGTCGCGATAAATTTGGAACCA	2811
Db	2230	AGAAAGCCGAGGACGGTTACAGGCCTTACAGAAAAAGGCTCGCGATAATTCGGAGGAA	2289
QY	2812	ACTCCGTTGATGGAAGGTGAGGGGAGATCTTTCAGAGTACTGGGTTTCAACCAGAGTCAA	2871
Db	2290	ATACCTTTGATGGAAGGCGAAGGGAGATATCTCATGGTGTGGGATTTAACGAGACTGAG	2349
QY	2872	AGGGCCATTTTGTACAGACTTTGATGAGGTATGGAGCTGGCAATTTTGATTGGAAGGAG	2931
Db	2350	AGGGATATATTTTACGTACGTTTAAGAGGTATGGAATCTTATTTTGAACACATTTGCC	2409
QY	2932	TTTGTTCCTCGCTTAAAGCAGAAAGACCTTTTGAAGAAATAAATGAATATGGAATACTCTTC	2991
Db	2410	GAAAACCTACCGATAATCTACAAATTTTAAAGTTATTACGCAATGGT-----	2459
QY	2992	TTGAAGCACATTGCTGAAGAAATFAGACGAGAAATCTCCAACTTTTCAGATGGTGTGCC	3051
Db	2460	-----CTATGCAGATGGAGTTCCC	2478
QY	3052	AAGGAAGGACTTAGAATAGAAAGATGTTCTAGTCAGAAATGCTCTTCTGATACTAGTTTCAG	3111
Db	2479	AAAGAAGGGATAAGCAGTGACGAGCTACTAGTGAGCATGACTTTTATGATGCTAGTAAAG	2538
QY	3112	GAGAAGTGAAATTTGTAGAAGATCATCCAGGGAACCTGTTTTTCCCCTCTCGCATTCIT	3171
Db	2539	GAGAAGTGTCATATTTTGGACACACCATCCGACAGCACCTGTITT---CTCTAACTACGTC	2595
QY	3172	GAAAGATCCCCGGA CTGAGAAGTGGAAAAAATTTGGAAGGAGGAACATGACAAAGATAATG	3231
Db	2596	ATCAGTAAGTACAATTTGAGAAATGAGCATTCTTAAGGAAGAACATGATAGGATACTG	2655
QY	3232	ATACGTGCTGTTTAAAGCATGGGTACGGACGGTGGCAAGCTATTGTTGATGACAAAGAG	3291
Db	2656	ATTCTGCTGTTTCCAAGCATGGCTATGGTAGATGGTGGCCATCGTTGAAGACGAAGAG	2715
QY	3292	TTGGGGATCCAAGAGCTTATCTGCAAAAGAAATTGAATTTCCCT	3333
Db	2716	ATAGGGTTCGAAGAGGTTGCCTGCAAAAGACCTGAATATCCCT	2757

RESULT 4

US-09-938-842A-2526  
; Sequence 2526, Application US/09938842A  
; Publication No. US20040009476A9  
; GENERAL INFORMATION:  
; APPLICANT: Harper, Jeff  
; APPLICANT: Kreps, Joel  
; APPLICANT: Wang, Xun  
; APPLICANT: Zhu, Tong  
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
; TITLE OF INVENTION: SAME, AND METHODS OF USE  
; FILE REFERENCE: SRIPI1300-3

; CURRENT APPLICATION NUMBER: US/09/938,842A			; CURRENT FILING DATE: 2001-08-24		
; PRIOR APPLICATION NUMBER: US 60/227,866			; PRIOR FILING DATE: 2000-08-24		
; PRIOR APPLICATION NUMBER: US 60/264,647			; PRIOR FILING DATE: 2001-01-16		
; PRIOR APPLICATION NUMBER: US 60/300,111			; PRIOR FILING DATE: 2001-06-22		
; NUMBER OF SEQ ID NOS: 5379			; LENGTH: 3204		
; TYPE: DNA			; ORGANISM: Arabidopsis thaliana		
US-09-938-842A-2526			Query Match 20.4%; Score 850.8; DB 11; Length 3204;		
			Best Local Similarity 61.0%; Pred. No. 1.9e-216;		
			Matches 1892; Conservative 0; Mismatches 697; Indels 513; Gaps 11;		
QY	280	GTTAGTCCTCTTAACGAGATAGATAAGATATTGGATTGTGAAATGCGTCCTACAAAATCT	339		
Db	121	GATTGTCCTCTTGGCGAGATTGAAAAGATTTGGACCGTGAATGGCGTCTACTGCATCT	180		
QY	340	AGTGAACAAGTTCTCCGAT---GCGGAACCGAAGCCAAATTTTGTGAAACAGTATCTC	396		
Db	181	AACAATCCGAATCTTCTGACAACGGGACACCAACCCTAGTTGTGCTGAAACAATATCTT	240		
QY	397	GTGAAGTGGAAAGGATTATCATACCTTCACCTGCTCTTGGGTGCTCGAGAAAGGATTCCAG	456		
Db	241	GTGAAGTGGAAAGGTTATCGTACTTGCACCTGCTCTTGGGTGCCGAGCAGGAGTTCGAG	300		
QY	457	AAGGCTTATAAGTCAAATCATCGTT-----TAAAAACCGAGTGAAACAAATTTTCACCGT	510		
Db	301	AAGGCTTACAAGTCCCATCTCTCATTTTAAAACTAAAAATTAAGGGTGAACGAGTTTAA	360		
QY	511	CAAAATGAGTCATTTCA-----ATAACAGCGGAAGATGATTTTGTGCCATACGTCCTGAG	564		
Db	361	GCTATGGACGTATTTCATCGCTGAGAACGGAGCTCATGAGTTTATCGCCATTCTGTCCTGAA	420		
QY	565	TGGACCACTGTTGATCGGATTCTTGCTGCAGAGAGGAAGATGGGAGCTGGAATATCTT	624		
Db	421	TGGAAAACTGTTGACCGGATTATTGCTTGCCGGGAAGGAGATGATGGGAGGAGTATCTG	480		
QY	625	GTCAAATATAAAGAGCTATCCTATGATGAATGTTATTGGGAGTCAGAAATCAGACATCTCA	684		
Db	481	GTGAAGTACAAAGAACTTTCTTACAGAAATAGTTACTGGGAATCTGAATCTGACATCTCA	540		
QY	685	ACCTTCCAGAAATGAAATTCAAAGGTTCAAGGATGTAAATCTAGAACTCGCAGAAAGTAAA	744		
Db	541	GATTTCCAGAAATGAAATCCAAAGGTTTAAAGGATATAAACTCTAGCTCTCGTAGAGATAA	600		
QY	745	GATGTTACCATAAAGAAATCCAGAGACTTTCAACAGTTTGTATCATACTCCTGAATTC	804		
Db	601	TATGTCGAAAATGAGAGAAATCGAGAGGAATTCAAACAGTTTGATCTTACTCCTGAATTT	660		
QY	805	CTCAAAGGTTGTATCATCCATACCAGTTGAGGGACTTAAATTTTGGCGTTCTCGTGG	864		
Db	661	CTTACAGGCACATTGCATACATACCAGCTCGAAGGGCTGAATTTCTTAAGGTATTTCATGG	720		
QY	865	TCAAACAGACCGCATGTAATCCTTGCTGATGAAATGGGACTAGGCAAGACAATTCAAAGC	924		
Db	721	TCCAAGAAAACCAATGTAATCCTTTCGGATGAAATGGGACTAGGCAAAACTATTTCAGAGC	780		
QY	925	ATTGCCCTTTTAGCTTCACTTTTGTAGGAGAACCTCATTCCTCCGATTTGGTAATTCCTCT	984		
Db	781	ATTGCCCTTCTTGGCCTCTTTGTTTCGAGGAAAACTTATCTCCTCATCTAGTTGTGCTCCT	840		
QY	985	CTATCGACTCTGCGTAACTGGGAGAGAGAGTTTGGCACATGGGCCCCACAGATGAACGTG	1044		
Db	841	CTCTCAACCATCCGTAACCTGGGAGAGGGAATTTGCCACCTGGGCACCAATATGAATGTG	900		
QY	1045	GTTATGTAATTTGGCACTGCGCAAGCTCGAGCAGTTATCAGAGAACATGAGTTTACTTAA	1104		

Db	901	GTTATGTACACTGGGGATTTCGGAAGCACGTGATGTTTATATGGGAACATGAGTTCTACTTC	960		
QY	1105	TCGAAAGATCAAAAAAAGATCAAGAAAAAGAAATCTGGACAAATAAGTAGCGAAAGCAAG	1164		
Db	961	TCTGAAGGGCGTAAAGAGTTAAAG-----	984		
QY	1165	CAAAAAAGAATCAAGTTTGATGTCTCCTCACATCGTATGAGATGATCAACCTAGATTCA	1224		
Db	985	-----	984		
QY	1225	GCAGTTCTAAACCAATTAAGTGGGAGTGATGATTGTTGATGAAGGTATCGACTGAAA	1284		
Db	985	-----	984		
QY	1285	AATAAGGATTCAAAGCTGTCTCTTCATTGACACAGTATTAAGTAACCAACCGTATTCTT	1344		
Db	985	-----	984		
QY	1345	CTGACAGGAACACCACCTTCAGAACAACTTGGATGAACCTTTTCATGCTCATGATTTTCTT	1404		
Db	985	-----	984		
QY	1405	GATCGGGGAAGTTTGGRAAGTTTGGAGGAGTTCCAGGAGAGTTCAAAGATATTAATCAA	1464		
Db	985	-----	984		
QY	1465	GAGGAGCAGATCTCAAGGTTGCACAAAATGTTGGCTCCACATTTGCTCAGAAGGGTAAAA	1524		
Db	985	-----	987		
QY	1525	AAAGACGTAATGAAAGACATGCCCCCAAAAAAGGAGCTCATTTTTCGCGTGTGATCTGAGC	1584		
Db	988	GACGTGCTCAAGGATAAGGTGCCCCCAAGAGGAGCTCATTTTTCGCGTGTGATATGAGC	1647		
QY	1585	AGTCTGCAGAAAGAATATTACAAAAGCTATTTTACCCTTAATTATCAAGTATTGACAAAA	1644		
Db	1048	AGCCAGCAAAAGAAGTGTAACAAGCTGTTATTACCATAATTAATCAAGTCTTGACAAAG	1107		
QY	1645	AAGGAGGTGCTCAAATTTCCCTTAATAACATTTATGATGGAATTAAGAAAAGTATGCTGC	1704		
Db	1108	AAACGAGATGCTAAAATATC-----TAACGTGCTAATGAAATTAACCCAAAGTTTGCCT	1161		
QY	1705	CATCCTTATATGCTAGAGGTGTTGAGCCAGTTATTACGACGCAAAATGAAGCTTTTCAA	1764		
Db	1162	CATCCGTATCTGTTACCAGACTTTGAGCCAAAGATTTGAAGACGCAAAATGAAGCTTTTACA	1221		
QY	1765	CAACTTTTGGAGTCTTGTGAAAGCTGCAACTTTCTAGATAAAATGATGGTCAAACTGAAA	1824		
Db	1222	AAACTATTGGAAGCTTCGGAAGTTTGCAACTGTTAGACAAAATGATGGTCAAACTCAA	1281		
QY	1825	GAGCAAGGACACAGAGTCTCTTAATATACACACAGTTTCAGCATATGCTGGACTTACTTGA	1884		
Db	1282	GAGCAGGGGCACAGAGTGTCTAATCTATACCAATTCACACACACACCTTTTACCTTTTGGAA	1341		
QY	1885	GACTACTGTACCCATAAGAAATGGCAGTACGAGCGGAATTTGATGGAAAGGTGGCGGAGCT	1944		
Db	1342	GACTACTTCACTTTCAAGAATTTGAAATTAAGAACGGATTGATGGAAAGATCAGTGGACCT	1401		
QY	1945	GAGCGGCAAAATACGCATAGATCGGTTCAATGCCAAAAAATTTCTAACAAAGTTTGTGTTTG	2004		
Db	1402	GAGAGGCAAGTAAGATAGATCGGTTTAATGACGAAAAAATCTCGAACCGGTTCTGTTTGTG	1461		
QY	2005	CTCTCCAAGAGCTGGTGGCTTAGGAATAAAATCTTGCAACGGCTGATACAGTAATCAT	2064		
Db	1462	CTCAGTACAAGAGCTGGTGAATAGGGATAAACTTGGCAACAGCTGATACAGTCACTCAT	1521		
QY	2065	TATGACAGTACTGGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCAFCGACTT	2124		
Db	1522	TATGACAGTATTGGAATCCTCATGCTGATCTCCAGGCTATGGCTAGAGTTTCAATCGGCTT	1581		
QY	2125	GGCCAAACAAATAAGGTGATGATTTATAGGCTCATAAACCGAGGCCACCATTTGAAGAAAGG	2184		
Db	1582	GGCCAAACAAATAAGGTAATGATCTACAGACTGATTTCACAAAGGTACTGTAGAGAACGG	1641		

QY	2185	ATGATGCAATTGACTAAAGAAAGAAATGGTTCTAGAGCATCTTGTGTGGAAACTCAA	2244
Db	1642	ATGATGGAGATAACCAAGAACAAAGATGCTTCTAGAGCATCTCGTCGTGGG-----A	1692
QY	2245	ACACAAAACATTAATCAGGAAGAGTTAGATGACATCATCAGGTATGGATCAAGGAGCTT	2304
Db	1693	AAACAACATCTTTGTCAGGACGAGTTAGATGACATCATCAAATATGGTCCAAGGAACTT	1752
QY	2305	TTTGCTAGTGAAGATGATGAAGCAGGAAAGTCTGGAAAAAATTCATTATGATGCGGCT	2364
Db	1753	TTCTCTGAAGAGATGATGAAGCAGGAGGTCTGGAAAAATCCATTACGATGCTGCT	1812
QY	2365	ATAGACAAATTCGTTGATCGTGATCTCGTGGAGGCAGAGGAAGTCTCAGTGGATGAA	2424
Db	1813	ATCGAACAGTTGCTAGATCGTAATCATGTAGATGCTGTGGAGGTCTCACTGGATGAG	1872
QY	2425	GAGGAGAAATGGATTCTTAAAGGCTTTCAAAGGTGGCTAATTTTGAATATATAGATAAAAT	2484
Db	1873	GAGGAAACTGATTTTAAAGAAATTTCAAGGTGGCTAGTTTGGAGTATGTAGATGAA	1932
QY	2485	GAGGCAGCAGCATTAGAGGCCACAGAGAGTCTGCTGCTGAAAGCAAATCTTCAGCAGGCAAT	2544
Db	1933	AATGAGGCTGCAGCATTAGAGGAGGCCAAGCTATAGAAAACAATCTTCAGTAAGAAAT	1992
QY	2545	TCTGATAGAGCAAGTTATTGGGAAGAGTTGTTAAAAAGATAAAATTTGAGCTGCACCAGGCT	2604
Db	1993	GCTGATAGAACAAAGTCACTGGAAGATTGCTGAAAGACAAATATGAAGTGCACAACAGCT	2052
QY	2605	GAGGAGCTTAATGCTCTTGAAAAAGGAAGACAGATCGCAAGCAGTTTGGTATCCATTGAA	2664
Db	2053	GAGGAGCTCAGTCTCTTGGAAGAGAGGAAGAGAAAAACGGCAAGCAGGTGATGTA--TGCT	2109
QY	2665	GAAGATGATCTTGTGGTTTGGAAAGATGTGAGCTCTGATGGAGATGAAAGTTATGAAGCT	2724
Db	2110	GAAGATGATCTTGTGGTTTAGAAGAAATCTCTGATGAGGAGGACGAATATGTCTTGAT	2169
QY	2725	GAGTCAACA-----GATGGTGAAGCAGCAGGACAA	2754
Db	2170	GATTTAAAAGTGACCTCTGATGAGGAGGAGGAAGCAGATGAACCCGAAGCAGCTCGACAG	2229
QY	2755	GGAGTTC--AGACGGGTCGACGGCCGTACAGAGAAAGGGTCGCGATAATTGGAAACCA	2811
Db	2230	AGAAAGCCGAGGACGGTTACAAGGCCTTACAGAAAAAGGGCTCGCGATAATTCCGAGGAA	2289
QY	2812	ACTCCGTTGATGGAAGGTGAGGGGAGATCTTTCAGAGTACTGGGTTTCAACCAGAGTCAA	2871
Db	2290	ATACCTTGTAGGAAGCGGAAGGGAGATATCTCATGGTCTGGGATTTAAACGAGACTGAG	2349
QY	2872	AGGGCCATTTTGTACAGACTTTTGATGAGGTATGGAGCTGGCAATTTTGATTGGAAGGAG	2931
Db	2350	AGGATATATTTTACGTACGTTTAAGAGGTATGGAATCTTATTTTGAACACATTGCC	2409
QY	2932	TTTGTTCCTCGCTTAAAGCAGAAGACCTTTTGAAGAAATAAATGAATATGGAATACTCTTC	2991
Db	2410	GAAAAACCTACCGGATAATTCTACAAATTTTAAAGTTATTACGGCAATGGT-----	2459
QY	2992	TTGAAGCACATTGCTGAAGAAATAGACGAGAAATCTCCAACCTTTTCAGATGGTGTGCCC	3051
Db	2460	-----CTATGCAGATGGAGTTCCC	2478
QY	3052	AAGGAAGGACTTAGAATAGAAGATGTTCTAGTCAGAAATGCTCTTCTGATFAGTTCAG	3111
Db	2479	AAAGAAGGGATAAGCAGTGACGAGCTACTAGTGAGCATGACTTTTATGATGCTAGTAAAG	2538
QY	3112	GAGAAGGTGAAATTTGTAGAAGATCATCCAGGGAACCTGTTTTCCCTCTCGCATTCCT	3171
Db	2539	GAGAAGTGTCAATTTTGGACAACCATCCGACAGCACCTGTTTT--CTCTAACTACGTC	2595
QY	3172	GAAAGATTTCCCGGACTGAGAAAGTGGAAAAATTTGGAAGGAGGAACATGACAAGATAATG	3231
Db	2596	ATCAGTAAGTACAATTTGAGAAATGGAGCAATTTTCTAAGGAAGAACATGATAGGATACTG	2655

QY	3232	ATACGTGCTGTTTTTAAAGCATGGGTACGGACGGTGGCAAGCTATTGTTGATGACAAAAGAG	3291
Db	2656	ATTCTCTGCTGTTTCCAAGCATGGCTATGGTAGATGGGTGGCCATCGTTGAAGACGAAGAG	2715
QY	3292	TTGGGATCCAAGAGACTTATCTGCAAAAGAAATTGAATTTCCCT	3333
Db	2716	ATAGGTTCCAAGAGGTTGCCTGCAAAAGACCTGAATATCCCT	2757
RESULT 5			
US-10-424-599-91500			
; Sequence 91500, Application US/10424599			
; Publication No. US20040031072A1			
; GENERAL INFORMATION:			
; APPLICANT: La Rosa Thomas J			
; APPLICANT: Kovalic David K			
; APPLICANT: Zhou Yihua			
; APPLICANT: Cao Yongwei			
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With			
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement			
; FILE REFERENCE: 38-21(53223)B			
; CURRENT APPLICATION NUMBER: US/10/424,599			
; CURRENT FILING DATE: 2003-04-28			
; NUMBER OF SEQ ID NOS: 285684			
; SEQ ID NO 91500			
; LENGTH: 2578			
; TYPE: DNA			
; ORGANISM: Glycine max			
; FEATURE:			
; NAME/KEY: unsure			
; LOCATION: (1)..(2578)			
; OTHER INFORMATION: unsure at all n locations			
; FEATURE:			
; OTHER INFORMATION: Clone ID: PAT_MRT3847_53638C.1			
US-10-424-599-91500			
Query Match 10.4%; Score 434.2; DB 13; Length 2578;			
Best Local Similarity 64.0%; Pred. No. 7.4e-105;			
Matches 729; Conservative 0; Mismatches 388; Indels 22; Gaps 4;			
QY	2258	ATCAGGAAGAGTTAGATGACATCATCAGGTATGGATCAAGGAGCTTTTGTAGTGAAG	2317
Db	20	AACAGGAAGAGTTGGATGACATCATAAAGATATGGCTCAAAGGAGTTATTTCAGATGAGA	79
QY	2318	ATGATGAAGCAGGAAAGTCTGAAAAAATTCATTATGATGATCGGCTATAGACAAATTGC	2377
Db	80	ATGATGAAGCAGGCAAAATCCCGCCAAATTCATTATGATGCTGCTGCTATTGATAGATTGC	139
QY	2378	TTGATCGTGATCTCGTGGAGGCAGAGGAAGTCTCAGTGGATGATGAAGGAGGAATGGAT	2437
Db	140	TGGATCGTGATCAAGTTGGAGATGAAGAGGCTACTTTGGATGATGAAGATGAAGATGGAT	199
QY	2438	TCTTAAAGGCTTCAAGGTGGCTAAATTTTGAATATATAGATGAAATGAGGCAGCAGCAT	2497
Db	200	TTCTGAAGGCTTTTAAGGTTGCAAAATTTTGAATATGTTGATGAGGCTGAGGCTGCAGCAG	259
QY	2498	TAGAGGCACAGAGAGTCTGCTGCTGAAAGCAAAATCTTCAGCAGGCAATTCGTATAGAGCAA	2557
Db	260	AGGAGGCAGCACAAAAAAGAGC-----AATGGAGACACTTAAACAGTTCAGAAAGGACAC	313
QY	2558	GTATTGCGGAAGAGTTGTTAAAAGATAAAATTTGAGCTGCACCAGGCTGAGGAGCTTAATG	2617
Db	314	ATTACTGGGAGGAGTTGTTAAGAGACAAGTATCAAGAGCATAAAGTTGAGGAGTTTAATG	373
QY	2618	CTCTTGGAAAAAGGAAGAGAGTTCGCAAGCAGTTGGTATCCATTGAAGAGATGATCTTG	2677
Db	374	CCCTGGGCAAGGGGAAGCGAAACCGGAAGTTGATGGTTCCGTGGGAGGATGACCTTG	433
QY	2678	CTGGTTTGGAAAGATGTGAGCTCTGATGG---AGATGAAAGTTATGAAGCTGAGTCAACAG	2734
Db	434	CTGGTCTGGAAGATGTAAGCTCTGATGCTGAAGATGACAAATTATGAAGCAGAGCTTACTG	493
QY	2735	ATGGTGAAGCAGCAGGACGAAGGAGTTTCAGACGGGTTCAGACGGCTCGACGGCCGTACAGAAAGAAA-----	2789



Db 494 ATGGTGATTCAAAATTCACACTGGAACACAACTAGAACTGCTAGAAAGACCTTATAAAAACAAGCTC 553  
QY 2790 -----GGGTCGGGATAAATTTGGAACCAACTCCGTTGATGGAAGGTGAGGGGAGATCTTTC 2844  
Db 554 TTACCGCCTCGGGATAGCACAGAGCCACTTCCTCTGATGGAAGGTGAAGAAAGCATTC 613  
QY 2845 AGAGTACTGGGTTTCAACCAGAGTCAAAGGGCCATTTTGTACAGACTTTGATGAGGTAT 2904  
Db 614 AGAGTCCTTGGTTTAAATCAAAATCAGAGGGCTCATTTGTGCAAATTTTGATGAGGTTT 673  
QY 2905 GGAGCTGGCAATTTGATTTGGAAGGAGTTTGTCTCGCTTAAAGCAGAAGACCTTTTGA 2964  
Db 674 GGGGTTGGTGATTTGATTTGGAAGGAGTTTACTTCCCGCATGAAACAGAAGACTTATGAA 733  
QY 2965 GAAATAAATGAATATGAATACTCTTCTTGAAGCACATTTGCTGAAGAAATAGACGAGAAT 3024  
Db 734 GAAATCAAAGACTATGGAACCCCTTTCTTGTCTCATATTGCTGAAGATATAACTG--AT 790  
QY 3025 TCTCCAAACCTTTTCAGATGGTGTGCCCAAGGAAGGACTTAGAATAGAAGATGTTCTAGTC 3084  
Db 791 TCCGCTACATTCACAGATGGTGTTCCAAAGAAGGACTCCGAATCCAAGATGTAATTGTT 850  
QY 3085 AGAATTGCTCTTCTGATACTAGTTTCAGGAGAGGTGAAAATTTGTAGAAGATCATCCAGGG 3144  
Db 851 AGGATTGCAGTCTGCTCTTGATAAGGGACAAAGGTGAAGTTTGTGTCAACAACATCCTCAA 910  
QY 3145 AAACCTGTTTTCCCTCTCGCATTTCTTGAAGATTCCTCCCGACTGAGAAGTGGAATAAT 3204  
Db 911 ACTCCATAATTTTCAGATGATATATTATTACGAAGCGAGTAATCTTGCCTGCCATGATT 970  
QY 3205 TGGAAGGAGGAACATGACAAGATAATGATACGTGCTGTTTTTAAAGCATGGGTACGGACGG 3264  
Db 971 TGGAGGACGCTGCCTCCCTAATGATTGTCTCCACTCCGCATGTACAATCAATAGCTTCAG 1030  
QY 3265 TGGCAAGCTATTGTTGATGACAAAGAGTTGGGGATCCAAAGAGCTTATCTGCAAGAAGATTG 3324  
Db 1031 TTGAATTTTCAGCTGAATTCGAATGGGATGAAGATCGAGTAGGTATATGCCCCCTAGTTG 1090  
QY 3325 AATTTCCCTCACATAAGTTTGTCTGCTGCTGCAACAAGCTGGTTTGCAGGGGCGAGAATGG 3383  
Db 1091 AATCTTTCTTAAATAAACTGACCAAAATACACGATGAATTAATTTCCAGCCTCATTTTGG 1149

RESULT 6  
US-09-960-253-142  
; Sequence 142, Application US/09960253  
; Patent No. US20020123619A1  
; GENERAL INFORMATION:  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Lodes, Michael J.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE REFERENCE: 210121.556  
; CURRENT APPLICATION NUMBER: US/09/960,253  
; CURRENT FILING DATE: 2001-09-20  
; NUMBER OF SEQ ID NOS: 187  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 142  
; LENGTH: 5739  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-960-253-142

Query Match 10.1%; Score 421.4; DB 9; Length 5739;  
Best Local Similarity 51.9%; Pred. No. 3.4e-101;  
Matches 1236; Conservative 0; Mismatches 1076; Indels 69; Gaps 10;  
QY 818 TACATCCATACCAAGCTTGAGGGACTTAATTTTTTTCGGGTTCTCGTGGTCAAAACAGACGC 877  
Db 2177 TGCACCCCTATCAAAATGAGGGCCTGAATTGGTTGCGCTTCTCCTGGGCTCAGGGCACTG 2236

QY 878 ATGTAATCCTTGTGATGAAATGGGACTAGGCAAGACAATTCAAAGCATTTGCCCTTTTAG 937  
Db 2237 ACACCATCTTGGCTGATGAGATGGGCTTTGGAAAACTGTACAGACAGCAGTCTTCTCTGT 2296  
QY 938 CTTCACTTTTGGAGGAG-----AACCTCATTCGCGCATTTGGTAATTGCTCCTCTATCGA 991  
Db 2297 ATTCCCTTTTACAAGGAGGGTCAITTCCAAAGGCCCTTCTCCTAGTAGCGGCCCTCTTTCTA 2356  
QY 992 CTCTGCGTAACCTGGGAGAGAGAGTTTGCCACACATGGGCCCCACAGATGAACGTGTTATGT 1051  
Db 2357 CCATCATCAACTGGGAGCGGGAGTTTGAATGTGGGCTCCAGACATGTATGTCGTAACCT 2416  
QY 1052 ATTTTGGCACTGCGCAAGCTCGAGCAGTTATCAGAGAACATGAGTTTTACTTATCGAAAG 1111  
Db 2417 ATGTGGGTGACAAAGACAGCCGCTGCCATCATCCGAGAGAAATGAGTTCTCTTTGAAGACA 2476  
QY 1112 ATCAAAAAAAGATCAAGAAAAAGAAATCTGGACAAATAAGTAGGAAAAAGCAAAAAA 1171  
Db 2477 ATGCCATTCGTGGTGGCAAGAGGCCTC-----CCGCATGAAGAAAGAGGCAT 2524  
QY 1172 GAATCAAGTTTGATGTCTCTCCTCACATCGTATGAGATGATCAACCTAGATTTCAGCAGTTC 1231  
Db 2525 CTGTGAATTCATGTCTGTGCTGACATCCTATGAATTGATCACCATTGACATGGCTATTT 2584  
QY 1232 TAAACCAATTAAGTGGGAGTGCATGATTGTTGATGAAGGTCTCGACTGAAAAATAAGG 1291  
Db 2585 TGGGCTCTATTGATTGGGCCCTGCCTCATCGTGGATGAAGCCCATCGGCTGAAGACAATC 2644  
QY 1292 ATTCAAAGCTGTTCTCTTCATTGACACAGTATTCAAGTAACCCGCTATTCTTCTGACAG 1351  
Db 2645 AGTCTAAGTTCTTCCGGGTATTGAATGGTTACTCACTCCAGCACAAAGCTGTTGCTGACTG 2704  
QY 1352 GAACACCACCTTCAGAACAACTTGGATGAACCTTTTCATGCTCATGCAATTTCTTGATGCGG 1411  
Db 2705 GGACACCATTACAAAACAATCTGGAAGAGTTGTTTCATCTGCTCAAATTTCTCACCCCG 2764  
QY 1412 GGAAGTTTGAAGTTTGGAGGAGTTCAGAGGAGGAGTTCAAAGATATTAAATCAAGAGGAGC 1471  
Db 2765 AGAGGTTCCACAATTTGGAAGGTTTTTTGGAGGAGTTTGTCTGACATTGCCAAGGAGACC 2824  
QY 1472 AGATCTCAAGGTTGCACAAAATGTGGCTCCACATTTGCTCAGAAGGGTAAAAAAGACG 1531  
Db 2825 AGATAAAAAAAACTGCATGACATGCTGGGGCCGCACATGTTGGCGGCTCAAAGCCGATG 2884  
QY 1532 TAATGAAAGACATGCCCCCCTCAAAAGGAGCTCATTTTGGTGTGTGATCTGAGCAGTCTGC 1591  
Db 2885 TGTTCAAGAACATGCCCTCCACAGACAGAACTAATTGTGCGTGTGAGCTGAGCCCTATGC 2944  
QY 1592 AGAAAGAATATTACAAAGCTATTTTACCCGTAATTATCAAGTATTGA---CAAAAAAGG 1648  
Db 2945 AGAAGAAATACTACAAGTACATCCTCACTCGAAATTTTGAAGCACTCAATGCCCGAGGTG 3004  
QY 1649 GAGGTGCTCAAATTTCCCTTAAATAACATTTATGATGGAATTACGAAAAAGTATGCTGCCATC 1708  
Db 3005 GTGGCAACCAGGTGTCTCTGCTGAATGTGGTATGGATCTTTAAGAAGTGTGCAACCATC 3064  
QY 1709 CTTA-----TATGCTAGAGGGTGTGAGCCAGTTTATTCACGACGCAATGAAGCTT 1759  
Db 3065 CATACCTCTTCCCTGTGGCTGCAATGGAAGCTCCTAAGATGCTAATGGCATGTATGATG 3124  
QY 1760 TCAACAACACTTTTGGAGTCTTGTGGAAGCTGCAACTTCTAGATAAAATGATGGTCAAAAC 1819  
Db 3125 GCAGTGCCCTAATCAGAGCATCTGGGAAAATATTGCTGCTGCAGAAAAATGCTCAAGAACC 3184  
QY 1820 TGAAGAGCAAGGACACAGAGTCCCTAATAATATACACACAGTTTCAGCATATGCTGGACTTAC 1879  
Db 3185 TTAAGGAGGGTGGGCATCGTGTACTCATCTTTTCCCAGATGATCAAGATGCTAGACCTGC 3244  
QY 1880 TTGAAGACTACTGTACCCATAGAATAATGGCAGTACGAGCGGAATTGATGGAAGGTTGGCG 1939  
Db 3245 TAGAGGATTTCTTGGAACATGAAGTTTATAAATACGAACGTCATGATGGTGAATCACTG 3304  
QY 1940 GAGCTGAGCGGCAAAATACGCATAGATCGGTTCAATGCCAAAAAATTTCTAACAAGTTTGT 1999

Db 3305 GGAACATGGCGCAAGAGCCATTGACCGCTTCAATGCACCGGTGCTCAGCAGTTCTGCT 3364  
Qy 2000 TTTTGTCTCCACAAGAGCTGGTGGCTTAGGAATAAATCTTGCAACGGCTGATACAGTAA 2059  
Db 3365 TCTTGTCTTCCACTCGAGCTGGGGCCCTTGGAATCAATCTGGCCACTGTCACACAGTTA 3424  
Qy 2060 TCATTTATGACAGTGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC 2119  
Db 3425 TTATCTATGACTCTGACTGGAACCCCAATAATGACATTCAGGCCCTTTAGCAGAGCTCACC 3484  
Qy 2120 GACTTGGCCAAACRAATAAGGTGATGATTTATAGGCTCATAAACCGAGGCACCATTTGAAG 2179  
Db 3485 GGATTGGGCAAAATAAAAAAGGTAATGATCTACCGGTTTGTACCCGTCGCTCAGTGGAGG 3544  
Qy 2180 AAAGGATGATGCAATTGACTAAAAAGAAAATGGTTCTAGAGCATCTTGTGTTGG----- 2234  
Db 3545 AGCGCATCAGCAGGTGGCAAGAAGAAAATGATGCTGACGCATCTAGTGGTGGCGCCTG 3604  
Qy 2235 -----GAAACTCAAAACACAAAACATTAATCAGGAAGAGTTAGATGACATCATCAGGTATG 2290  
Db 3605 GGCTGGGCTCAAAGACTGGATCTATGTCCAACAGGAGCTTGATGATATCCTCAAATTTG 3664  
Qy 2291 GATCAAAGGAGCTTTTGTGCTAGTGAAGATGATGAAGCAGGAAAGTCTGGAAAAATTCATT 2350  
Db 3665 GCACTGAGGAACTATTCAAGGATGAAGCCACTGATGGAGGAGGAGACAAACAAGAGGGAG 3724  
Qy 2351 ATGATGATCGGCTATAGACAAAATTGCTTGA-----TCGTGATCTCGTGAGGCAGAGG 2404  
Db 3725 AAGATAGCAGTGTATTCCACTACGATGATGAAGGCCATTGAACGGCTGCTAGACCGTAACC 3784  
Qy 2405 AAGTCTCAGTGGATGAAGAGGAGAAATGGATTCTTAAAGGCTTTCAAAGTGGCTAATT 2464  
Db 3785 AGGATGAGACTGAAGACACAGAAATTGCAGGGCATGAATGAATATTTGAGCTCATTCAAAG 3844  
Qy 2465 TTGAATATATAGATGAAAATGAGGCAGCAGCATTAGAGGCACAGAGAGTCGCTGCTGAAA 2524  
Db 3845 TGGCCCAGTATGTGTTACGGGAAGAAGAAATGGGGAGGAAGAGGAGGTAGAACGGGAAA 3904  
Qy 2525 GCAAATCTTCAGCAGGCAATTCTGATAGAGCAAGTTATTGGGAAGAGTTGTTAAAGATA 2584  
Db 3905 TCATTAAACAGGAAGAAAGTGTGGATCCTGA---CTACTGGGAGAAATTGCTGCGGCACC 3961  
Qy 2585 AATTTGAGCTGCACCGGCTGAGGAGCTTAAATGCTCTTGGAAAAAGGAAGAGATCGCA 2644  
Db 3962 ATTATGAGCAGCAGCAAGAAGATCTAGCCCCGAAATCTGGGCAAGGAAAAGAAATCCGTA 4021  
Qy 2645 AGCAGTTGGTAICCAATTGAAG---AAGATGATCTTGCTGGTTTGGAAGATGTGAGCTCTG 2701  
Db 4022 AACAGGTCAACTACAATGATGGCTCCCAGGAGGACCGAGATTGGCAGGACGACCACTCCG 4081  
Qy 2702 ATGGAGATGAAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGACAAGGAGTTC 2761  
Db 4082 ACAACCACTCCGATTACTCAGTGGCTTCAGAGGAAGGTGATGAAGACTTTTGATGAACGTT 4141  
Qy 2762 AGACGGGTGACCGGCCGTACAGAAAGAAAGGTCGCGATAAATTTGGAACCAA----- 2812  
Db 4142 CAGAAAGCTCCCCGTAGGCCAGTCGTAAGGGCCCTGCGGAATGATAAAGATAAGCCATTGC 4201  
Qy 2813 CTCGGTTGATGGAAGGTGAGGGGAGATCTTTACAGTACTGGGTTTCAACCAGAGTCAAA 2872  
Db 4202 CTCCTCTGTTGGCCCGTGTGGTGGGAATATTGAAGTACTTGGTTTTTAATGCTCGTCAGC 4261  
Qy 2873 GGGCCATTTTGTACAGACTTTTGATGAGGTATGG-----AGCTGGCAATTTTGATT 2923  
Db 4262 GAAAAGCCTTCTTAAATGCAATATATCGATATGATGATGTCACCTCAGGATGCTTTTACTA 4321  
Qy 2924 GGAAGGAGTTTGTTCCTCGCTTAAAGCAGAAAGACCTTTGAAGAAAATAAATGAATATGGAA 2983  
Db 4322 CCCAGTGGCTTGTAAAGACACCTGCGAGGCAAAATCAGAGAAAAGAGTTCAAGGCATATGCT 4381  
Qy 2984 TACTCTTCTTGAAGCACAATTGCTGAAGAAATAGACGAGAATTCTCCAACCTTTTCAGATG 3043

Db 4382 CTCTTTTCATCGGCATTTATGTGAGCCGGGGCAGATGGGGCTGAGACCTTTTGCTGATG 4441  
Qy 3044 GTGTGCCCAAGGAAGGACTTAGAATAGAAGATGTTTCTAGTCAGAAFTGCTTTCTGATAC 3103  
Db 4442 GTGTCCCCGAGAAGGCCTGTCTCGCCAGCATGTCTTACTAGAATTGGTGTATGTCTT 4501  
Qy 3104 TAGTTCAGGAGAAAGGTGAAATTTGTAGAAGATCATCCAGGG 3144  
Db 4502 TGATTCCGAAGAAGTTTCAGGAGTTTGAACATGTTAATGGG 4542  
RESULT 7  
US-09-962-436-288  
; Sequence 288, Application US/09962436  
; Patent No. US20020081301A1  
; GENERAL INFORMATION:  
; APPLICANT: Soppet, Daniel  
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu  
; TITLE OF INVENTION: Sets  
; FILE REFERENCE: 689290-75  
; CURRENT APPLICATION NUMBER: US/09/962,436  
; CURRENT FILING DATE: 2001-09-25  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,082  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/234,924  
; NUMBER OF SEQ ID NOS: 568  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 288  
; LENGTH: 6417  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-962-436-288  
Query Match 10.1%; Score 421.4; DB 9; Length 6417;  
Best Local Similarity 51.9%; Pred. No. 3.7e-101;  
Matches 1236; Conservative 0; Mismatches 1076; Indels 69; Gaps 10;  
Qy 818 TACATCCATACCAGCTTGAGGGACTTAATTTTTCGGTTCTCGTGGTCAAAAACAGACGC 877  
Db 2266 TGCACCCCTATCAAATGGAGGCCCTGAATTGTTGCGTTCTCCTGGECTCAGGGCACTG 2325  
Qy 878 ATGTAATCCTTGCTGATGAAATGGGACTAGGCAAGACAATTCAAAGCATTGCCCTTTTAG 937  
Db 2326 ACACCATCTTGGCTGATGAGATGGGCTTGGGAAAACGTACAGACAGCAGTCTTCCCTGT 2385  
Qy 938 CTTCACTTTTGGAGGAG-----AACCTCATTCGCGCATTTTGGTAATTGCTCCTCTATCGA 991  
Db 2386 ATTCCCTTTACAAGGAGGGTCATTCCAAAGGCCCTTCCCTAGTAGCGGCCCTCTTTCTA 2445  
Qy 992 CTCTGCGTAACTGGGAGAGAGAGTTTGCCACATGGGCCCCACAGATGAACGTGGTTATGT 1051  
Db 2446 CCATCATCAACTGGGAGCGGGAGTTTGAAATGTGGGCTCCAGACATGTATGTCGTAACCT 2505  
Qy 1052 ATTTTGGCACTGCGCAAGCTCGAGCAGTTATCAGAGAACATGATGTTTACTTATCGAAAG 1111  
Db 2506 ATGTGGGTGACAAGGACAGCCGTGCCATCATCCGAGAGAAATGAGTTCTCCTTTGAAGACA 2565  
Qy 1112 ATCAAAAAAAGATCAAGAAAAAGAAATCTGGACAAATAAGTAGCGAAAGCAAGCAAAAA 1171  
Db 2566 ATGCCATTTCGTGGTGGCAAGAGGCCCTC-----CCGCATGAAGAAAAGAGGCAT 2613  
Qy 1172 GAATCAAGTTTGATGTCCTCCTCACATCGTATGAGATGATCAACCTAGATTTCAGCAGTTC 1231  
Db 2614 CTGTGAAATTCATGTGCTGTGACATCCTATGAATTGATCACCATTGACATGGCTATTT 2673  
Qy 1232 TAAACCAATTAAAGTGGGAGTGCATGATTGTTGATGAAGGTTCATCGACTGAAAAATAAGG 1291  
Db 2674 TGGGCTCTATTGATTGGGCCCTGCCTCATCGTGGATGAAGCCCATCGGCTGAAGACAATC 2733  
Qy 1292 ATTCAAAGCTGTTCTCTTCATTGACACAGTATTCAAGTAACCAACCGTATTCTTCTGACAG 1351  
Db 2734 AGTCTAAGTTCTTCCGGGTATTGAATGGTTACTCACTCCAGCACAAAGCTGTTGCTGACTG 2793





i ORGANISM: Homo sapiens									
US-10-096-534-28									
Query Match		10.1%;	Score 421.4;	DB 15;	Length 6417;				
Best Local Similarity		51.9%;	Pred. No. 3.7e-101;						
Matches 1236;		Conservative 0;	Mismatches 1076;	Indels 69;	Gaps 10;				
QY	818	TACATCCATACCAGCTTGAGGGACTTAATTTTTTGGGGTTCTCGTGGTCAAAACAGACGC	877						
Db	2266	TGCACCCCTATCAAAATGGAGGGCCTGAATTGGTTGCGCTTCTCCTGGGCTCAGGGCACTG	2325						
QY	878	ATGTAATCCTTGCTGATGAAATGGGACTAGGCAAGACAATAATTCAAAGCAATTGCCCTTTTAG	937						
Db	2326	ACACCATCTTGGCTGATGAGATGGGCCCTTGGGAAAACTGTACAGACAGCAGTCTTCCTGT	2385						
QY	938	CTTCACCTTTTGGAGGAG-----AACCTCATTCGCGCATTTGGTAATTGCTCCTCTATCGA	991						
Db	2386	ATTCCCTTTACAAGGAGGGTCATTCCAAAGGCCCTTCCTAGTGAGCGGCCCTCTTTCTA	2445						
QY	992	CTCTGCGTAACCTGGGAGAGAGAGTTTGGCCACATGGGCCCCACAGATGAACGTGGTTATGT	1051						
Db	2446	CCATCATCAACTGGGAGCGGGAGTTTGAAATGTGGGCTCCAGACATGTATGTGCTAACCT	2505						
QY	1052	ATTTTGGCACTGCGCAAGCTCGAGCAGTTATCAGAGAACATGAGTTTTTACTTATCGAAAG	1111						
Db	2506	ATGTGGGTGACAAGGACAGCCGTGCCATCATCCGAGAGAATGAGTTCTCCTTTGAAGACA	2565						
QY	1112	ATCAAAAAAAGATCAAGAAAAAGAAATCTGGACAATAAGTAGCGAAAGCAAGCAAAAAA	1171						
Db	2566	ATGCCATTTCGTGGTGGCAAGAGGCCCTC-----CCGCATGAAGAAAGAGGCAT	2613						
QY	1172	GAATCAAGTTTGTATGTCCTCCTCACATCGTATGAGATGATCAACCTAGATTTCAGCAGTTC	1231						
Db	2614	CTGTGAATTCATGTGCTGCTGACATCCTATGAATGTATCACCATTGACATGGCTATTT	2673						
QY	1232	TAAAACCAATTAACTGGGAGTGCATGATTGTTGTGAAGGTTCATCGACTGAAAAATAAGG	1291						
Db	2674	TGGGCTCTATTGATTGGGCCCTGCCCTCATCGTGGATGAAGCCCATCGGCTGAAGAACAAATC	2733						
QY	1292	ATTCAAAGCTGTTCTCTTCATTGACACAGTATTCAAAGTAACCAACCGTATTCCTCTGACAG	1351						
Db	2734	AGTCTAAGTCTTCCGGGTATTGAATGGTTACTCACTCCAGCACAAAGCTGTTGCTGACTG	2793						
QY	1352	GAAACACCATTTCAGAACAACTTGGATGAACCTTTTCATGCTCATGCAATTTTCTTGATGCGG	1411						
Db	2794	GGACACCATTACAAAAACAATCTGGAAGAGTTGTTTCATCTGCTCAAATTTCTCACCCCCG	2853						
QY	1412	GGAAGTTTGGNAGTTTGGAGGAGTTCAGGAGGAGTTCAAAGATATTAAATCAAGAGGAGC	1471						
Db	2854	AGAGGTTCCACAATTTGGAAGGTTTTTTTGGAGGAGTTTGCTGACATTGCCAAGGAGGACC	2913						
QY	1472	AGATCTCAAGGTTGCACAAAATGTTGGCTCCACATTTGCTCAGAAGGGTAAAAAAGACG	1531						
Db	2914	AGATAAAAAAACTGCATGACATGCTGGGCCGCGCACATGTTGCGCGGCTCAAAGCCGATG	2973						
QY	1532	TAATGAAAGACATGCCCCCCTCAAAAAGGAGCTCATTTTGGCGTGTGATCTGAGCAGTCTGC	1591						
Db	2974	TGTTCAAGAACATGCCCTCCAAGACAGAACTAATTGTGCGTGTGGAGCTGAGCCCTATGC	3033						
QY	1592	AGAAAGAATATTACAAAGCTATTTTACCCTGTAATTATCAAGTATTGA---CAAAAAAGG	1648						
Db	3034	AGAAGAAATACTACAAGTACATCCTCACTCGAAATTTTGAAGCACTCAATGCCCGAGGTG	3093						
QY	1649	GAGGTGCTCAAAATTTCCCTTAATAAACATTATGATGGAATTACGAAAAAGTATGTCGCCATC	1708						
Db	3094	GTGGCAACCAGGTGTCCTGCTGAATGTGGTGATGGATCTTAAGAAGTCTGCAACCAATC	3153						
QY	1709	CTTA-----TATGCTAGAGGGTGTGAGCCAGTTATTTCAGCAGCGCAATGAAGCTT	1759						
Db	3154	CATACCTCTTCCCTGTGGCTGCAATGGAAGCTCCTAAGATGCCCTAATGGCATGTATGATG	3213						
QY	1760	TCAAACAACTTTTGGAGTCTTGTGGAAAGCTGCAACTTCTAGATAAAATGATGTCCTCAAC	1819						

Db	3214	GCAGTGCCTTAATCAGAGCATCTGGGAAATATTATTGCTGTCAGAAAAATGCTCAAGAAC	3273						
QY	1820	TGAAAGAGCAAGGACACAGAGTCTCTAATATATACACACAGTTTCAGCATATGCTGGACTTAC	1879						
Db	3274	TTAAGGAGGGTGGGCATCGTGTACTCATCTTTTCCAGATGATCAAGATGCTAGACCTGC	3333						
QY	1880	TTGAAGACTACTGTACCCATAAGAAATGGCAGTACGAGCGAAATGATGGAAAGGTTGGCG	1939						
Db	3334	TAGAGGATTTCTTGGAAACATGAAGGTTATAAATACGAACGCATCGATGGTGAATCACTG	3393						
QY	1940	GAGCTGACGGCAAAATACGCATAGATCGGTTCAATGCCAAAAATTTAAACAAGTTTGT	1999						
Db	3394	GGAACATGCGGCAAGAGGCCATTGACCGCTTCAATGCACCGGGTGTCTAGCAGTTCTGCT	3453						
QY	2000	TTTTTGTCTCCACAAGAGCTGGTGGCTTAGGAATAAATCTTGCACGGCTGATACAGTAA	2059						
Db	3454	TCTTGCTTTCACCTCGAGCTGGGGCCTTGAATCAATCTGGCCACTGCTGACACAGTTA	3513						
QY	2060	TCATTTATGACAGTGACTGGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC	2119						
Db	3514	TTATCTATGACTCTGACTGGAACCCCAATAANTGACATTCAGGCCTTTAGCAGAGCTCAC	3573						
QY	2120	GACTTGGCCAAAACAAATAAAGGTGATGATTTATAGGCTCATAAACGAGGCACCATTTGAAG	2179						
Db	3574	GGATTGGCAAAATAAAAGGTAATGATCTACCGGTTTGTGACCGCTGCTCAGTGGAGG	3633						
QY	2180	AAAGGATGATGCAATTGACTAAAAAGAAAAATGGTTCTAGAGCATCTTGTGTGG-----	2234						
Db	3634	AGCGCATCAGCAGGTGGCAAGAGAAATGATGCTGACGCATCTAGTGGTGGCGCCTG	3693						
QY	2235	----GAAACTCAAAACACAAAAACATTAATCAGGAAGAGTTAGATGACATCATCAGGTATG	2290						
Db	3694	GGCTGGGCTCCAAGACTGGATCTATGTCCAACACAGGAGCTTGATGATATCCTCAAAATTG	3753						
QY	2291	GATCAAAGGAGCTTTTGTCTAGTGAAGATGATGAAGCAGGAAGTCTTGGAAAAAATTCATT	2350						
Db	3754	GCACTGAGGAACATATTCAAGGATGAAGCCACTGATGGAGGAGAGACAACAAGAGGGAG	3813						
QY	2351	ATGATGATGCGGCTATAGACAAAATTGCTTGA-----TCGTGATCTCGTGGAGGCAGAGG	2404						
Db	3814	AAGATAGCAGTGTATTCCACTACGATGATAAGGCCATTGAAACGGCTGCTAGACCGTAACC	3873						
QY	2405	AAGTCTCAGTGGATGATGAAGAGGAGAATGGATTCTTAAAGGCTTTCAAAGGTGGCTAATT	2464						
Db	3874	AGGATGAGACTGAAGACACAGAAATTCAGGGCATGAATGAATATTTGAGCTCATTCAAAG	3933						
QY	2465	TTGAATATATAGATGAAAAATGAGGCAGCAGCATTAGAGGCACAGAGAGTCGCTGCTGAAA	2524						
Db	3934	TGGCCCAGTATGTGTA CGGGAAGAAGAAATGGGGAGGAAGAGGAGGTAGAACGGGAAA	3993						
QY	2525	GCAAACTTTCAGCAGGCAATTCTGTATAGACCAAGTTATTGGGAAGAGTTGTTAAAGATA	2584						
Db	3994	TCATTAAACAGGAAGAAAGTGTGGATCCTGA---CTACTGGGAGAAATTGTGCGGCACC	4050						
QY	2585	AATTGAGCTGCACCGCTGAGGAGCTTAATGCTCTTGGAAAAAGGAAGAGAGATCGCA	2644						
Db	4051	ATTATGAGCAGCAGCAAGAAGATCTAGCCGAAATCTGGGCAAGGAAAAAGAAATCCGTA	4110						
QY	2645	AGCAGTTGCTATCCATTGAAG---AAGATGATCTTGTGTTTGGAAAGATGTGAGCTCTG	2701						
Db	4111	AACAGGTCAACTACAATGATGGCTCCAGGAGGACCGAGATTGGCAGACGACCACTCCG	4170						
QY	2702	ATGGAGATGAAAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGACAAGGAGTTC	2761						
Db	4171	ACAACCAGTCCGATTACTCAGTGGCTTCAGAGGAAGGTGATGAAGACTTTGATGAACGTT	4230						
QY	2762	AGACGGTTCGACGGCCGTACAGAAGAAAGGTCGCGATAATTTGGAACCAA-----	2812						
Db	4231	CAGAAGCTCCCCGTAGGCCCACTCGTAAGGCCCTGCGGAATGATAAAGATAAGCCATTGC	4290						
QY	2813	CTCCGTTGATGGAAGGTGAGGGGAGATCTTTCAGAGTACTGGGTTTCAACCAGAGTCAAA	2872						
Db	4291	CTCCTCTGTTGGCCCGTGTGTTGGTGGGAATATTGAAGTACTTGGTTTTTAATGCTCGTCAGC	4350						



Db 3523 GGAACATGCGGCAAGAGGCCATTGACCGCTTCAATGCACCGGTGCTCAGCAGTTCTGTCT 3582

Qy 2000 TTTTGTCTCTCCACAAGAGCTGGTGGCTTAGGAATAAAATCTTGCAACGGTGTATACAGTAA 2059

Db 3583 TCTTGTCTTCCACTCGAGCTGGGGCCCTTGGAAATCAATCTGGCCACTGCTGACACAGTTA 3642

Qy 2060 TCATTTATGACAGTGACTGGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC 2119

Db 3643 TTATCTATGACTCTGACTGGAACCCCCATAATGACATTCAGGCCCTTTAGCAGAGCTCACC 3702

Qy 2120 GACTTGGCCAAACAAATAAGGTGATGATTTTATAGGCTCATAAACCGAGCCACCATTTGAAG 2179

Db 3703 GGATTGGGCAAAATAAAAAGGTAATGATCTACCGGTTTGTGACCCGTCAGTGGAGG 3762

Qy 2180 AAAGGATGATGCAATTGACTAAAAAGAAAATGGTTCTAGAGCATCTTGTGTTGG----- 2234

Db 3763 AGCGCATCACGCAAGTGGCAAGAAAGAAAATGATGCTGACGCATCTAGTGGTGGCCTG 3822

Qy 2235 -----GAAACTCAAAACACAAAAACATTAAATCAGGAAGAGTTAGATGACATCATCAGGTATG 2290

Db 3823 GGCTGGGCTCCAAGACTGGATCTATGTGCCAAACAGGAGCTTGTATGATATCTCAAATTTG 3882

Qy 2291 GATCAAAAGGAGCTTTTGTCTAGTGAAGATGATGAAGCAGGAAAGTCTCGAAAAAATTCATT 2350

Db 3883 GCACTGAGGAACATAAAGGATGAAGCCACTGATGGAGGAGACAAACAGGGAG 3942

Qy 2351 ATGATGATCGGGCTATAGACAAATTGCTTGA-----TCGTGATCTCGTGGAGGCAGAGG 2404

Db 3943 AAGATAGCAGTGTATTCCACTACGATGATAAGGCCATTGAACGGCTGTAGACCGTAACC 4002

Qy 2405 AAGTCTCAGTGGATGAAGAGGAGGAATGGATTCTTAAAGGCTTTCAAGGTGGCTAAATT 2464

Db 4003 AGGATGAGACTGAAGACACAGAAATTGCAGGGCATGAATGAATAATTGAGCTCATTTCAAAG 4062

Qy 2465 TTGAAATATATAGATGAAAAATGAGGCAGCAGCATATTAGAGGCACAGAGATCGCTGATAA 2524

Db 4063 TGGCCCAGTATGTGTACGGGAAGAAAGAAATGGGGAGGAAGAGGATGAGAACGGGAAA 4122

Qy 2525 GCAAAATCTTCAGCAGGCAATTCATGATAGAGCAAGTTATTGGGAAGAGTTGTTAAAAAGATA 2584

Db 4123 TCATTAAACAGGAAGAAAGTGTGGATCCTGA---CTACTGGGAGAAATTGCTGCGGCACC 4179

Qy 2585 AATTGAGCTGCACCAGSCTGAGGAGCTTAATGCTCTTGAAAAAAGGAAGAGTCGCA 2644

Db 4180 ATTATGAGCAGCAGCAAGAAGATCTAGCCCCGAATCTGGGCAAGGAAGAAATCCGTA 4239

Qy 2645 AGCAGTTGGTATCCATTGAAG-- --AAGATGATCTTGCTGGTTTGGGAAGATGTGAGCTCTG 2701

Db 4240 AACAGGTCAACTACAATGATGGCTCCCAGGAGGACCGAGATTGGCAGGACGACCATCCG 4299

Qy 2702 ATGGAGATGAAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGACAAGGAGTTC 2761

Db 4300 ACAACCAGTCCGATTACTCAGTGGCTTCAGAGGAAGGTGATGAAGACTTTTGATGAACGTT 4359

Qy 2762 AGACGGGTGACGGCCGTACAGAAAGAAAGGGTCGCGATAAATTTGGAACCAA----- 2812

Db 4360 CAGAAAGCTCCCCGTAGGCCCAGTCGTAAAGGGCTGCGGAATGATAAAGATAAAGCCATTGC 4419

Qy 2813 CTCGGTTGATGGAAGGTGAGGGGAGATCTTTTCAGAGTACTGGGTTTCAACCCAGAGTCAAA 2872

Db 4420 CTCCTCTGTTGGCCCCGTGTGGTGGGAATATTGAAGTACTTGGTTTTAATGTCGTCAGC 4479

Qy 2873 GGGCCATTTTGTACAGACTTTTGATGAGGTATGG-----AGCTGGCAATTTTGATT 2923

Db 4480 GAAAAGCCTTTCTTAATGCAATTATGCGATATGGTATGCCACCTCAGGATGCTTTTACTA 4539

Qy 2924 GGAAGGAGTTTGTTCCTCGCTTAAAGCAGAAGACCTTTGAAGAAATAAATGAATATGGAA 2983

Db 4540 CCCAGTGGCTTGTAAAGAGACTGCGAGGCCAAATCAGAGAAAGAGTTCAAAGGCATATGTCT 4599

Qy 2984 TACTCTTCTTGAAGCACATGTGCTGAAGAAAATAGACAGAAATTCCTCAAACCTTTTCAGATG 3043

Db 4600 CTCTTTTCATGCGGCATTTATGTGAGCCGGGCGAGATGGGGCTGAGACCTTTGCTGATG 4659

Qy 3044 GTGTGCCCAAGGAAGGACTTAGAATAGAAAGATGTTCTAGTCAGAATTGCTCTTCTGATAC 3103

Db 4660 GTGTCCCCCGAGAAGGCCTGTCTCGCCAGCATGTCTTACTAGAAATGGTGTATGTCTT 4719

Qy 3104 TAGTTCAGGAGAAAGGTGAAATTTGTAGAAGATCATCCAGGG 3144

Db 4720 TGATTCGAAGAAGGTTTCAGGAGTTTGAACATGTTAATGGG 4760

RESULT 10

US-10-117-722-325

; Sequence 325, Application US/10117722

; Publication No. US20030219744A1

; GENERAL INFORMATION:

; APPLICANT: Tang, Y. Tom

; APPLICANT: Liu, Chenghua

; APPLICANT: Asundi, Vinod

; APPLICANT: Zhang, Jie

; APPLICANT: Drmanac, Radoje T.

; TITLE OF INVENTION: No. US20030219744A1el Nucleic Acids and

; TITLE OF INVENTION: Polypeptides

; FILE REFERENCE: 784CIP2BCIP

; CURRENT APPLICATION NUMBER: US/10/117,722

; CURRENT FILING DATE: 2002-04-04

; PRIOR APPLICATION NUMBER: 09/620,312

; PRIOR FILING DATE: 2000-07-19

; PRIOR APPLICATION NUMBER: 09/552,317

; PRIOR FILING DATE: 2000-04-25

; PRIOR APPLICATION NUMBER: 09/488,725

; PRIOR FILING DATE: 2000-01-21

; NUMBER OF SEQ ID NOS: 1104

; SOFTWARE: pt\_FL\_genes Version 1.0

; SEQ ID NO 325

; LENGTH: 6475

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (219)..(5957)

US-10-117-722-325

Query Match 10.1%; Score 421.4; DB 16; Length 6475;

Best Local Similarity 51.9%; Pred. No. 3.7e-101;

Matches 1236; Conservative 0; Mismatches 1076; Indels 69; Gaps 10;

Qy 818 TACATCCATACCAGCTTGAGGGACTTAATTTTTCGGGTTCTCGTGGTCAAAACAGACGC 877

Db 2395 TGCACCCCTATCAAAATGGAGGGCCTGAATGGTTGCGCTTCTCTCGGCTCAGGGCACTG 2454

Qy 878 ATGTAATCCTTGCTGATGAAATGGGACTAGGCAAGACAATTCAAAGCATTGCCCTTTTAG 937

Db 2455 ACACCATCTTGGCTGATGAGATGGCCTTGGGAAAACTGTACAGACAGCAGTCTTCTGT 2514

Qy 938 CTTCACTTTTGGAGGAG-----AACCTCATTCGGCATTTTGGTAATTGCTCCTCTATCGA 991

Db 2515 ATTCCCTTTACAAAGGAGGGTCATTCCAAAGGCCCTTCTCTAGTAGCGCCCTCTTTCTA 2574

Qy 992 CTCTGCGTAACACTGGGAGAGAGATTGTCACATGGGCCCCACAGATGAACGTGGTTATGT 1051

Db 2575 CCATCATCAACTGGGAGCGGGAGTTTGAATGTGGCTCCAGACATGTATGTCGTAACCT 2634

Qy 1052 ATTTTGGCACTGCGCAAGCTCGAGCAGTTATCAGAGAACATGAGTTTACTTATCGAAAG 1111

Db 2635 ATGTGGGTGACAAGGACAGCCGTGCCATCATCCGAGAGAAATGAGTTCCTTTGAAGACA 2694

Qy 1112 ATCAAAAAAAGATCAAGAAAAAGAAATCTGGACAAATAAGTAGCGGAAAGCAAGCAAAAA 1171

Db 2695 ATGCCATTCGTGGTGGCAAGAGGCCCTC-----CCGCATGAAGAAAGAGGCAT 2742

Qy 1172 GAATCAAGTTTGATGTCTCCTCACATCGTATGAGATGATCAACCTAGATTCAGCAGTTC 1231

Db 2743 CTGTGAAATTCATGTGCTGTGCTGACATCTCTATGAATTGATCAACCATTGACATGGCTATTT 2802













Db 3461 AGAACACCAGTCCAAAGTTTTCAGGGTTCCTCAATGGTTACAAGTAGATCATAAGTTGC 3402  
QY 1343 TTCTGACAGGAACACCACCTTCAGAACAACTTCGATGAACCTTTTCATGCTCATGCTTTTC 1402  
Db 3401 TGCTGACAGGAACCCCATTCGAGAATAATCTGGAGGAGCTCTTCCATCTCTGAACTTCC 3342  
QY 1403 TTGATGCGGGGAAGTTTGGAAAGTTTGGAGGAGTTCAGGAGGAGTTCAAAGATATTAATC 1462  
Db 3341 TCACCCAGAGAGATTAAACAATTGGAGGGTTCCTGGAGGAGTTTGCTGACATATCCA 3282  
QY 1463 AAGAGGAGCAGATCTCAAGGTTTGACAAAATGTTGGCTCCACATTTGCTCAGAAGGTTAA 1522  
Db 3281 AAGAGGACCAGATCAAGAAACTGCATGATTTGCTGGGGCCACACATGCTCGGAGACTCA 3222  
QY 1523 AAAAAACGTAATGAAGACATGCCCCCAAAAAAGGAGCTCATTTTGGCTGTTGATCTGA 1582  
Db 3221 AGGCAGATGCTTTAAGAACATGCCAGCCCAAGACAGAGCTCATCGTTCGGTGGAGCTAA 3162  
QY 1583 GCAGTCTGCAGAAAGAATATTACAAAGCTATTTTACCCGTAATTATCAAGTATTGA --- 1639  
Db 3161 GCCCATGCAGAAAGAAATACTACAAATACCTGACTCGAAATTTTGGGCCTTGAATT 3102  
QY 1640 CAAAAAAGGAGGTGCTCAAAATTTCCCTTAATAACATTATGATGAATTACGAAAAGTAT 1699  
Db 3101 CACGAGGTGTGGAAACCAGGTGTCGCTGCTTAATATCATGATGGAATCTTAAGAAGTGCT 3042  
QY 1700 GCTGCCATCCTATATGCTAGAGGGTGTTG-----AGCCAGTTATTACGACGCAA 1750  
Db 3041 GCAACCATCCATACCTTTTCCCGTGGCTGATGGAGTCCCCCAAACTCCCAGTGGGG 2982  
QY 1751 ATGAAGCTTTCAAACAACCTTTTGGAGTCTTGTGGAAGCTGCAACTTCTAGATAAAATGA 1810  
Db 2981 CTTATGAGGTGGGCACCTTATTAAGTCGTCTGGGAAGCTCATGCTGCTCCAGAAGATGC 2922  
QY 1811 TGGTCAAACCTGAAGAGCAAGGACACAGAGTCTCTAATAATACACACAGTTTTCAGCATATGC 1870  
Db 2921 TGCGAAGCTGAAGAGCAAGGACACCGAGTGTCTCATCTTCTCGCAGATGACCAAAATGT 2862  
QY 1871 TGGACTTACTTGAAGACTACTGTACCCATAAGAAATGGCAGTACGAGCGAAATTGATGGAA 1930  
Db 2861 TAGACTTGCTTGAGGACTTCTTAGACTATGAAGGCTACAAGTATGAGCGCATCGATGGTG 2802  
QY 1931 AGGTTGGCGGAGCTGAGCGGCAAAATACGCATAGATCGGTTCAATGCCAAAAATTTCTAACA 1990  
Db 2801 GTATCACGGGTGCCCTGAGGCAGGAGGCCATCGATCGGTTTAAATGCTCTGGGGCCCAAC 2742  
QY 1991 AGTTTGTGTTTGTCTCTCCACAAGAGCTGGTGGCTTAGGAATAAATCTTGCAACGGCTG 2050  
Db 2741 AATTCTGCTTCCTCTGTCACCCGAGCTGGGGCCCTGGGCATCAATCTGGCCACTGCTG 2682  
QY 2051 ATACAGTAATCATTTATGACAGTGACTGGAATCCTCATGCTGATCTTCAAGCAATGGCTA 2110  
Db 2681 ACATGTCTCATCTTTGATTCTGACTGGAACCCCAATAATGACATCCAGGCCTTTTAGCC 2622  
QY 2111 GAGCTCATCGACTGGCCAAAACAAATAAGGTGATGATTTATAGGCTCATATAAACCGAGGCA 2170  
Db 2621 GGGCTCATCGGATTTGGCCAGGCCAACAAAGTGAATGATTTTACCGGTTTGTGACTCGCGCGT 2562  
QY 2171 CCATTGAAGAAAGGATGATGCAATTGACTTAAAAAGAAAATGGTTCTAGAGCATCTTGTGTTG 2230  
Db 2561 CAGTGAAGAGCGGAATCACACAAGTGGCCAAGAGAAAGATGATGCTGACACACCTGGTTG 2502  
QY 2231 TTGG-----GAAACTCAAACACAAAAACATTAAATCAGGAAGAGTTAGATGACATCA 2281  
Db 2501 TCGCGCCTGGGCTGGGCTCCAAGGCAGGCTCCATGTCCAAGCAGGAGCTTGACGACATTC 2442  
QY 2282 TCAGGTATGGATCAAAGGAGCTTTTGTGCTAGTGAAGATGATGAAGCAGGAAAAGTCTG --- 2338  
Db 2441 TCAAATTTGGACTGAAGAGCTATTCAAGGATGAAAAACGAGGGGAGAAACAAGGAGGAGG 2382  
QY 2339 -----GAAAAATTCAATATGATGATGCGGCTATAGACAAATTTGCTTGATCTCG 2392

Db 2381 ACAGCAGTGTGATTATTATGACAATGAGGCCATCGCTCGGCTGTTGGACCGGAACCAGG 2322  
QY 2393 TGGAGGCAGAGGAAGTCTCAGTGGATGATGAAGAGGAGAAATGGATTCTTAAAGGCTTTTCA 2452  
Db 2321 ATGCAACTGAGGA-----CACTGACGTGCAGAACATGAATGAGTATCTCAGCTCCTTCA 2268  
QY 2453 AGGTGGCTAAATTTTGAATATATAGATGAAAATGAGGCAGCAGCATTAGAGGCACAGAGAG 2512  
Db 2267 AGGTGGCACAGTACGTCGTGCGGGAAGAAGACAAGATTGAGGAAATTGACGAGAGATCA 2208  
QY 2513 TCGCTGCTGAAAGCAAATCTTCAGCAGGCAATTTCTGATAGACCAAGTTATTGGGAAGAGT 2572  
Db 2207 TCAAGCAGGAGGAGAAATGTGGACCTTGAC-----TACTGGGAGAAGC 2166  
QY 2573 TGTTAAAGATAAATTTTGAAGTTCACCCAGGCTGAGGAGCTTAATGCTCTTGGAAAAAGGA 2632  
Db 2165 TGCTGAGGCATCACTATGAGCAACAGCAGGAAGACCTAGCCCGGAATCTAGGCAAGGGCA 2106  
QY 2633 AGAAGAGTCGCAAGCAGTTGGTATCCATTGAAGAAGATGATCTTGTGTTTGGAAAGATG 2692  
Db 2105 AGCGGGTTCGCAAGCAAGTTAACTACAATGATGCTGCTCAGGAAGACCAAGACAACCAGT 2046  
QY 2693 TGAGTCTGATGGAGATGAAAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGAC 2752  
Db 2045 CAGAGTACTCGGTGGTTTCAGAGGAGGAGTGAAGACTTCGATGAACGCTCCTGAAGGGC 1986  
QY 2753 AAGGAGTTCAGACGGGTCGACGGCCGTACAGAAGAAAGGTCGCGATAAATTTGGAACCAA 2812  
Db 1985 GTAGACAGTCAAAGAGGACGCTCCGGAA-----TGAGAAAAGATAAGCCACTGC 1938  
QY 2813 CTCGTTGATGGAAGGTGAGGGGAGATCTTTTCAGAGTACTGGGTTTCAACCAGAGTCAA 2872  
Db 1937 CTCACACTGTGGCCCGAGTCGGGGGCAACATTGAGGTGCTGGGCTTCAACACCCGTCAGC 1878  
QY 2873 GGGCCATTTTGTACAGACTTTTGATGAGGTATGG-----AGCTGGCAATTTTGATT 2923  
Db 1877 GGAAGGCTTTCCTCAATGCTGTGATCGCTGGGGGATGCCACCACAGGATGCCCTTCACCA 1818  
QY 2924 GGAAGGAGTTTGTTCCTCGCTTAAAGCAGAAAGACCTTTTGAAGAAAATAAATGAATATGAA 2983  
Db 1817 CACAGTGGCTGGTGGCGGACCTGAGGGGCAAGACTGAGAAGGAGTTTAAGGCCTATGTGT 1758  
QY 2984 TACTCTTCTTGAAGCACATTGCTGAAGAAAATAGACGAGAAATCTCCAACCTTTTCAGATG 3043  
Db 1757 CTTTGTTCATCGCCCATCTGTGTGAGCCTGGGGCAGACGGCTCTGAAACCTTTTGCCGATG 1698  
QY 3044 GTGTGCCCCAAGGAAGGACTTAGAATAGAAGATGTTCTAGTCAGAAATTTGCTCTCTGATAC 3103  
Db 1697 GGGTCCCTCGGAGGAGTGAAGTCCAGCAGAGGTGTTGACCCCGCATTTGGAGTCATGTCTC 1638  
QY 3104 TAGTTCAGGAGAAGGTGAA 3122  
Db 1637 TCGTCAAAAAGAAAGGTGCA 1619

RESULT 14  
US-10-044-090-370  
; Sequence 370, Application US/10044090  
; Publication No. US20020137081A1  
; GENERAL INFORMATION:  
; APPLICANT: Olga Bandman  
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION  
; FILE REFERENCE: PA-0028 US  
; CURRENT APPLICATION NUMBER: US/10/044,090  
; CURRENT FILING DATE: 2002-01-09  
; NUMBER OF SEQ ID NOS: 850  
; SOFTWARE: PERL Program  
; SEQ ID NO 370  
; LENGTH: 7805  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature

; OTHER INFORMATION: Incyte ID No. US20020137081A1 235885.5									
US-10-044-090-370									
Query Match 9.8%; Score 409.6; DB 14; Length 7805;									
Best Local Similarity 52.5%; Pred. No. 6.2e-98;									
Matches 1240; Conservative 0; Mismatches 1024; Indels 96; Gaps 12;									
QY	810	AGGCTTGTTACATCCATACACAGCTTGAGGGACTTAATTTTGGGGTTCCTCGTGGTCAAA	869						
Db	2692	AGGCACCCTGCACATGTATCAGTTGGAAGGGCTGAACCTGGCTACGCTTCTCCTGGGCCCA	2751						
QY	870	ACAGACGCATGTAATCCTTGCTGTGATGAAATGGGACTAGGCAAGACAAATTCAAAGCATTGC	929						
Db	2752	GGGCACTGACACCATTTCTAGCTGTATGAGATGGGGCTAGGCAAGACCACATACAAACCATCGT	2811						
QY	930	CCTTTTAGCTTCACTTTTGTGAGGAGAACCTCA-----TTCCGCATTTGGTAAATTGCTCC	983						
Db	2812	CTTCCTCTACTCACTCTACAAGGAGGGCCACACAAAAGGTCCCTTCCTGGTGAGTGCCCC	2871						
QY	984	TCTATCGACTCTCGGTAACTGGGAGAGAGAGTTTGCCACATGGGCCCCACACAGATGAACGT	1043						
Db	2872	ACTCTTACCATCATTAAGTGGAGCGGAGTTCAGATGTGGGCACCCAAATTCATGT	2931						
QY	1044	GGTTATGTATTTTGGCACTGCGCAAGCTCGAGCAGTTATCAGAGAACATCAGTTTACTTT	1103						
Db	2932	GGTGACATACACGGGTGACAAGGACAGCCGGGCCATCATTCGTGAGAATGAATTCCTCT	2991						
QY	1104	ATCGAAAGAT-CAAAAAAAGATCAAGAAAAAGAAATCTGGACAAATAAGTAGCGAAAGCA	1162						
Db	2992	TGAGGACAAATGCCATCAAGGGGGCAAGAAAGCTTTTAAAGATGAAGAGGGGCACAGG-	3050						
QY	1163	AGCAAAAAAGAATCAAGTTTGATGTCTCTCCATCAGATCGTATGAGATGATCAACCTAGATT	1222						
Db	3051	-----TGAAAGTTCATGTTCTCTGACATCGTATGAGCTGATCACCATTGATC	3098						
QY	1223	CAGCAGTTCTAAAACCAATTAAGTGGGAGTGCAATGTTGTGATGAAGGTCAATCGACTGA	1282						
Db	3099	AGGCAGCACTTGGTTCATCCGCTGGCCGTGCTTGTGGTAGATGAGGCCCATCGACTCA	3158						
QY	1283	AAAATAAGGATTCAAAGCTGTTCTCTTCATTTGACACAGTATTCAGATAACCAACCGTATTC	1342						
Db	3159	AGAACAAACCAGTCCAAGTTTTCAGGGTTCTCAATGGTTACAAGATAGATCATAAGTTGC	3218						
QY	1343	TTCTGACAGGAACACCACTTCAGAACAACTTGGATGAACCTTTTCATGCTCATGCAATTTTC	1402						
Db	3219	TGCTGACAGGAACCCCAATTGCAGAAATAATCTGGAGGAGCTCTTCCATCTCCTGAACTTC	3278						
QY	1403	TTGATCGGGGAAGTTTGGAAAGTTTGGAGGAGTTCAGGAGGAGTTCAAAGATATTAATC	1462						
Db	3279	TCACCCACAGAGAGATTTAAACAACTTGGAGGGCTTCCTGGAGGAGTTTGTGACATATCCA	3338						
QY	1463	AAGAGGAGCAGATCTCAAGGTTGCACAAAAATGTTGGCTCCACATTTGCTCAGAAGGGTAA	1522						
Db	3339	AAGAGGACCAGATCAAGAAACTGCATGATTTGCTGGGGCCACACATGCTGCGGAGACTCA	3398						
QY	1523	AAAAAGACGTAATGAAGACATGCCCCCCCCAAAAAGGAGCTCATTTTGCCTGTTGATCTGA	1582						
Db	3399	AGGCAGATGTCTTTAAGAAATGCCAGCCCAAGACAGAGCTCATCGTTTCGGGTGGAGCTAA	3458						
QY	1583	GCAGTCTGCAGAAAGAAATATTACAAAGCTATTTTACCCTTAATTATCAAGTATTGA---	1639						
Db	3459	GCCCCATGCAGAAGAAATACTACAAATACATCCTGACTCGAAATTTTGAGGCCCTTGAATT	3518						
QY	1640	CAAAAAAGGGAGGTGCTCAAAATTTCCCTTAATAACATATATGATGGAATTACGAAAAAGTAT	1699						
Db	3519	CACGAGGTGGTGGGAACCAAGTGTGCTGCTTAAATATCATGATGGATCTTAAAGAAGTGCT	3578						
QY	1700	GCTGCCATCCTTATATGCTAGAGGGTGTG-----AGCCAGTTATTACGACGCAA	1750						
Db	3579	GCAACCATCCATACCTTTTCCCGTGGCTGCTATGGAGTCCCCCAAACTCCCCAGTGGGG	3638						
QY	1751	ATGAAGCTTTCAAACCAACTTTTGGAGTCTTGTGGAAAGCTGCAACTTCTAGATAAAATGA	1810						

Db	3639	CTTATGAGGGTGGGGCACTTATTAAATCGTCTGGGAAGCTCATGCTCTCCAGAAGATGC	3698						
QY	1811	TGGTCAAAACTGAAAGAGCAAGGACACAGAGTCCTAATATATACACAGTTTCAGCATATGC	1870						
Db	3699	TGCGAAAGCTGAAGGAGCAAGGACACCGAGTGCTCATCTTCTCGAGATGACCAAATGT	3758						
QY	1871	TGGACTTACTTGAAGACTACTGTACCCATAAGAAATGGCAGTACGACGCAATTGATGGAA	1930						
Db	3759	TAGACTTGTCTGAGGACTTCTTAGACTATGAAGGCTACAAGTATGAGCGCATCGATGGTG	3818						
QY	1931	AGGTTGGCGGAGCTGAGCGGCAATFACGCATAGATCGGTTCAATGCCAAAAATTTCTAACA	1990						
Db	3819	GTATCACGGGTGCCCTGAGGCAGGAGGCCCATCGATCGGTTTAATGCTCTGGGCCCAAC	3878						
QY	1991	AGTTTGTGTTTTGCTCTCCACAAGAGCTGGTGGCTTAGGAATAAATCTTGCAACGGCTG	2050						
Db	3879	AATTCGTCTCCTCTGTCCACCCGAGCTGGGGCCTGGGCATCAATCTGGCCACTGCTG	3938						
QY	2051	ATACAGTAATCATTTATGACAGTGAAGTGAATCCTCATGCTGATCTTCAAGCAATGGCTA	2110						
Db	3939	ACACTGTATCATCTTTGATTTCTGACTGGAACCCCCATAATGACATCCAGGCCCTTAGCC	3998						
QY	2111	GAGCTCATC-GACTTGGCCAAACAAATAAGGTGATGATTTATAGGCTCATAAACCGAGGC	2169						
Db	3999	GGGCTCATCGGATTTGGCCAGGCCAACAAAGTGATGATTTACCGGTTTGTGACTCGCGCG	4058						
QY	2170	ACCATTGAAGAAAGGATGATGCAATTGACTAAAAAGAAAAATGGTTCTAGAGCATCTTGT	2229						
Db	4059	TCAGTGGAAAGAGCGAATCACACAAGTGGCCCAAGAGAAAGATGATGCTGACACACCTGGTT	4118						
QY	2230	GTGG-----GAAACTCAAAACACAAAAACATTAATCAGGAAGAGTTAGATGACATC	2280						
Db	4119	GTGCGGCTGGGCTGGCTCCAAAGCAGGCTCCATGTCCAAGCAGGAGCTTGACGACATT	4178						
QY	2281	ATCAGGTATGGATCAAAGGAGCTTTTTCGTAGTGAAGATGATGAAGCAGGAAAGTCTG--	2338						
Db	4179	CTCAAAATTTGGCACTGAAGAGCTATTCAAGGATGAAAAACGAGGGGAGAACAGGAGGAG	4238						
QY	2339	-----GAAAAATTCATTATGATGATGCGGCTATAGACAAATTCGTTGATCGTATCTC	2391						
Db	4239	GACAGCAGTGTGATTCATTATGCAATGAGGCCATCGCTCGGCTGTTGGACCGGAACAG	4298						
QY	2392	GTGAGGCGAGGAAAGTCTCAGTGGATGATGAAGAGGAGAAATGGAATTCCTTAAAGGCTTTC	2451						
Db	4299	GATGCAACTGAGGA-----CACTGACGTGCAGAACATGAATGAGTATCTCAGCTCCTTC	4352						
QY	2452	AAGGTGGCTAAATTTGAATATATAGATGAATAATGAGGCAGCAATFAGAGGCACAGAGA	2511						
Db	4353	AAGTGGCACAGTACGTGTCGTGCGGGAAGAAAGACAAGATTGAGGAAATGAGCGAGAGATC	4412						
QY	2512	GTGCTGCTGAAAAGCAAATCTTCAGCAGGCAATTCGTATAGAGCAAGTTATTGGGAAGAG	2571						
Db	4413	ATCAAGCAGGAGGAGAAATGTGGACCCCTGAC-----TACTGGGAGAAG	4454						
QY	2572	TTGTTAAAAGATAAAATTTGAGCTGCACCAGGCTGAGGAGCTTAATGCTCTTTGGAAAAAGG	2631						
Db	4455	CTGCTGAGGCATCACTATGAGCAACAGCAGGAAGACCTAGCCCGGAATCTAGGCAAGGGC	4514						
QY	2632	AAGAGAAAGTCGAAGCAGTTGGTATCCAATGAAGAAAGATGATCTTGTGGTTTGGAAAGAT	2691						
Db	4515	AAGCGGTTTCGCAAGCAAAGTTAACTACAATGATGCTGCTCAGGAAGACCAAGAACCCAG	4574						
QY	2692	GTGAGCTCTGATGGAGATGAAAGTTATGAAGCTGAGTCAACAGATGTTGAAAGCAGCAGGA	2751						
Db	4575	TCAGAGTACTCGTGGGTTTCAGAGGAGGAGGATGAAGACTTCGATGAACGCTCCTGAAGGG	4634						
QY	2752	CAAGGAGTTCAGACGGGTTCGACGGCCGTACAGAAAGAAAGGTCGCGATAAATTTGGAACCA	2811						
Db	4635	CGTAGACAGTCAAGAGGCGAGCTCCGGAA-----TGAGAAAGATAAGCCACTG	4682						
QY	2812	ACTCCGTTGATGGAAGGTGAGGGGAGATCTTTCAGAGTACTGGGTTTCAACCAGAGTCAA	2871						
Db	4683	CCTCCACTGCTGGCCCCAGTCGGGGGGCAACATTGAGGTGCTGGGCTTCAACACCCGTCAG	4742						





Db	4119	GTGCGGCTGGCTGGGCTCCAAGGCAGGCTCCATGTCCAAGCAGGAGCTTGACGACATT	4178
Qy	2281	ATCAGGTATGGATCAAAAGGAGCTTTTGTAGTGAAGATGATGAAGCAGGAAAAGTCTG--	2338
Db	4179	CTCAAATTTGGCACTGAAGAGCTATTCAAGGATGAAAAACGAGGGGAGAACAAAGGAGGAG	4238
Qy	2339	-----GAAAAATTCATTATGATGATCGGCTATAGACAAATTGCTTGATCGTGATCTC	2391
Db	4239	GACAGCAGTGTGATTCAATTATGACAATGAGGCCATCGCTCGGCTGTTGGACCGGAACCG	4298
Qy	2392	GTGGAGGCAGAGGAAGTCTCAGTGGATGATGAAGAGGAGAAATGGAATTCCTTAAAGGCTTTC	2451
Db	4299	GATGCAACTGAGGA-----CACTGACGTGCAGAACATGAATGAGTATCTCAGCTCCTTC	4352
Qy	2452	AAGGTGGCTAATTTGAATATATATAGATGAATAATGAGGCAGCAGCATTAGAGGCACAGAGA	2511
Db	4353	AAGGTGGCACAGTACGTCTGTCGGGAAGAACAAAGATTGAGGAAATTGAGCGAGAGATC	4412
Qy	2512	GTCGCTGCTGAAGCAAAATCTTCAGCAGGCAATCTGATAGAGCAAGTTATTTGGGAAGAG	2571
Db	4413	ATCAAGCAGGAGGAGAATGTGACCCTGAC-----TACTGGGAGAAG	4454
Qy	2572	TTGTTAAAAGATAAATTTGAGCTGCACCAGGCTGAGGAGCTTAATGCTCTTGGAAAAAGG	2631
Db	4455	CTGCTGAGGCATCACTATGAGCAACAGCAGGAAGACCTAGCCCCGGAATCTAGGCAAGGCC	4514
Qy	2632	AAGAGAACTCGCAAGCAGTTGGTATCCATTGAAGAAGATGATCTTGTGCTGGTTTGAAGAT	2691
Db	4515	AAGCGGGTTCGCAAGCAAGTTAACTACAATGATGCTGCAGGAAGAACCAAGACCAACAG	4574
Qy	2692	GTGAGCTCTGATGGAGATGAAAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGA	2751
Db	4575	TCAGAGTACTCGGTGGGTTCAGAGGAGGAGGATGAAGACTTCGATGAACCTCCTGAAGGG	4634
Qy	2752	CAAGGAGTTCAGACGGGTCGACGGCCGTACAGAAGAAAGGTCGCCGATAAATTTGGAACCA	2811
Db	4635	CGTAGACAGTCAAAAGAGGCAGCTCCGGAA-----TGAGAAAGATAAGCCACTG	4682
Qy	2812	ACTCCGTTGATGGAAGGTGAGGGGAGATCTTTCAGAGTACTGGGTTTCAACCAGAGTCAA	2871
Db	4683	CCTCCACTGCTGGCCGAGTCGGGGGCAACATTGAGGTGCTGGGCTTCAACACCCCGTCAG	4742
Qy	2872	AGGGCCATTTTGTACAGACTTTGATGAGGTATGG-----AGCTGGCAATTTTGAT	2922
Db	4743	CGGAAGGCTTTCCTCAATGCTGTGATGCGCTGGGGGATGCCACCACAGGATGCCTTCACC	4802
Qy	2923	TGGAAGGAGTTTGTTCCTCGCTTAAAGCAGAAAGACCTTTGAAGAAATAAATGAATATGGA	2982
Db	4803	ACACAGTGGCTGGTGGGGACCTGAGGGGCAAGACTGAGAAGGAGTTTAAAGCCCTATGTG	4862
Qy	2983	ATACTCTTCTTGAAGCACATTCGTGAAGAAATAGACGAGAAATTCCTCAACCTTTTCAGAT	3042
Db	4863	TCTTTGTTTCATGCGCCATCTGTGTGAGCCTGGGGCAGACGGCTCTGAAACCTTTGCCGAT	4922
Qy	3043	GGTGTGCCCCAAGGAAGACTTAGAATAGAAGATGTTCTAGTCAGAATTGCTCTTGATA	3102
Db	4923	GGGTCCCTCGGGAGGGACTGAGTCGCCAGCAGGAGTGTGACCCGCAATTGGAGTCATGTCT	4982
Qy	3103	CTAGTTCAGGAGAAGGTGAA	3122
Db	4983	CTCGTCAAAAAGAGGTGCA	5002

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 14:48:31; Search time 9894 Seconds  
(without alignments)  
12607.068 Million cell updates/sec

Title: US-10-049-137-1  
Perfect score: 4177  
Sequence: 1 atgagtagttagtgaggag.....caactggtaaatcaagattc 4177

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :		EST:*	
	1:	em_estba:*	
	2:	em_esthum:*	
	3:	em_estin:*	
	4:	em_estmu:*	
	5:	em_estov:*	
	6:	em_estpl:*	
	7:	em_estro:*	
	8:	em_htc:*	
	9:	gb_est1:*	
	10:	gb_est2:*	
	11:	gb_htc:*	
	12:	gb_est3:*	
	13:	gb_est4:*	
	14:	gb_est5:*	
	15:	em_estfun:*	
	16:	em_estom:*	
	17:	em_gss_hum:*	
	18:	em_gss_inv:*	
	19:	em_gss_pln:*	
	20:	em_gss_vrt:*	
	21:	em_gss_fun:*	
	22:	em_gss_mam:*	
	23:	em_gss_mus:*	
	24:	em_gss_pro:*	
	25:	em_gss_rod:*	
	26:	em_gss_phg:*	
	27:	em_gss_vrl:*	
	28:	gb_gss1:*	
	29:	gb_gss2:*	

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	910.2	21.8	939	14	CF651673
2	612.8	14.7	639	9	AU237917
3	528	12.6	528	9	AV551529
4	415.4	9.9	4226	11	BC059082

	5	408.2	9.8	3137	11	AK034549	AK034549 Mus muscu
	6	398.8	9.5	610	10	BE204297	BE204297 EST396973
	7	387.6	9.3	745	14	CB635984	CB635984 OSIEB16P
c	8	385	9.2	563	9	AI998628	AI998628 701546371
	9	373	8.9	597	10	BE204298	BE204298 EST396974
	10	354.6	8.5	589	13	BQ139596	BQ139596 NF021H12P
	11	351.4	8.4	583	10	BE205450	BE205450 EST398126
	12	336	8.0	778	14	CA240279	CA240279 SCSEFL406
	13	308	7.4	348	13	BQ834228	BQ834228 ALEST0190
c	14	300	7.2	503	28	B29916	B29916 T19D20TRB T
	15	298.2	7.1	673	14	CD812910	CD812910 BN10.022P
	16	296.2	7.1	4209	11	AK049686	AK049686 Mus muscu
	17	287.8	6.9	557	9	AU287846	AU287846 AU287846
	18	285.6	6.8	2740	11	AK039580	AK039580 Mus muscu
	19	280.2	6.7	492	9	AU225820	AU225820 AU225820
	20	278	6.7	843	14	CB667080	CB667080 OSJNE14E
	21	277.2	6.6	655	12	BI928046	BI928046 EST547935
	22	276.8	6.6	673	12	BM406422	BM406422 EST580737
	23	276	6.6	653	13	BU000377	BU000377 QGG24K16.
	24	272.4	6.5	680	13	BQ518530	BQ518530 EST625945
	25	267.4	6.4	583	9	AV835455	AV835455 AV835455
	26	265.6	6.4	656	14	CB079334	CB079334 hp69g07.b
	27	257.4	6.2	810	13	CA077816	CA077816 SCRFAM102
	28	253.4	6.1	1811	11	AK015218	AK015218 Mus muscu
	29	253.2	6.1	2877	11	AK030430	AK030430 Mus muscu
	30	251.6	6.0	3161	11	AK040495	AK040495 Mus muscu
c	31	251.4	6.0	830	13	BX748807	BX748807 BX748807
	32	242.8	5.8	3607	11	AK018129	AK018129 Mus muscu
	33	235.2	5.6	752	14	CB525972	CB525972 UI-M-FY0-
	34	233.8	5.6	502	12	BJ186506	BJ186506 BJ186506
	35	232.4	5.6	532	14	CA170938	CA170938 SCRUSB106
	36	232.4	5.6	874	13	BU194813	BU194813 AGENCOURT
	37	231	5.5	935	9	AL519189	AL519189 AL519189
	38	228.4	5.5	813	14	CA324135	CA324135 UI-M-FY0-
	39	227.8	5.5	767	14	CD349796	CD349796 UI-M-FY0-
	40	226.2	5.4	675	14	CA500458	CA500458 WHE4019 H
	41	225.8	5.4	775	14	CB525974	CB525974 UI-M-FY0-
c	42	223.8	5.4	818	14	CF784972	CF784972 AGENCOURT
	43	223.6	5.4	516	12	BI321298	BI321298 saf50b11.
	44	223.4	5.3	766	14	CB247361	CB247361 UI-M-FY0-
	45	221.4	5.3	2168	11	AK052092	AK052092 Mus muscu

ALIGNMENTS

RESULT 1	CF651673	939 bp	mRNA	linear	EST 06-NOV-2003
LOCUS	08-L020524-066-003-002-SP6P	MPIZ-ADIS-066	Arabidopsis thaliana	CDNA	
DEFINITION	clone MPIZp20010023Q	5-PRIME, mRNA	sequence.		
ACCESSION	CF651673				
VERSION	CF651673.1	GI:37427433			
KEYWORDS	EST.				
SOURCE	Arabidopsis thaliana (thale cress)				
ORGANISM	Arabidopsis thaliana				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
AUTHORS	1 (bases 1 to 939)				
TITLE	Schmid,K.J., Soerensen,T.R., Stracke,R., Torjek,O., Altmann,T., Mitchell-Olds,T. and Weisshaar,B.				
JOURNAL	Large-scale identification and analysis of genome-wide single-nucleotide polymorphisms for mapping in Arabidopsis thaliana				
MEDLINE	Genome Res. 13 (6), 1250-1257 (2003)				
PUBMED	22683290				
COMMENT	12799357				
	Contact: Weisshaar B				
	ADIS DNA core facility at MPIZ				
	Max-Planck-Institute for Plant Breeding Research				
	Carl-von-Linne Weg 10, 50829 Koeln, Germany				
	Fax: 00492215062851				
	Email: weisshaar@mpiz-koeln.mpg.de				



Insert Length: 939 Std Error: 0.00  
Plate: 3 row: 0 column: 02  
Seq primer: SP6P;  
Location/Qualifiers  
1. .939  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/cultivar="Wassilewskija-0 (Ws-0)"  
/db\_xref="GABI:938264"  
/db\_xref="taxon:3702"  
/clone="MPIZp2001O023Q"  
/tissue\_type="root"  
/lab\_host="E. coli TOP10"  
/clone\_lib="MPIZ-ADIS-066"  
/note="Vector: pCMVSPORT6; Site\_1: Sali; Site\_2: NotI;  
cDNA library from Arabidopsis thaliana, accession  
Wassilewskija-0; roots from three weeks old plants grown  
on MS-plates at 26M-OC with 16 hours light/day; library  
was made at the Max-Planck-Institute for Plant Breeding  
Research, Cologne, Germany; cloning sites Sali-NotI,  
primer sites and orientation:  
SP6-Sali-CCACGGCTCCG-5prime-cDNA-polyA-CC-NotI-T7; GATEWAY  
compatible; Note: Sequencing granted in the context of the  
GABI Arabidopsis Verbund I: Genetic Diversity,  
'Establishment of high-efficiency SNP-based mapping tools  
and development of methods for genome-wide mutation  
detection' PI: Bernd Weisshaar Sequence submission managed  
by RZPD/GABI-Primary database: http://gabi.rzpd.de This  
clone is available from RZPD; contact RZPD (clone@rzpd.de)  
for further information."

ORIGIN

Query Match 21.8%; Score 910.2; DB 14; Length 939;  
Best Local Similarity 99.5%; Pred. No. 7.5e-195;  
Matches 923; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 1 ATGAGTAGTTTGGTGAGAGGCTTCGCATACGATCTGATAGGAAACCAAGTTTATAACCTA 60  
Db 13 ATGAGTAGTTTGGTGAGAGGCTTCGCATACGATCTGATAGGAAACCAAGTTTATAACCTA 72  
QY 61 GATGATTCTGATGATGACGACTTCGTTCCCTAAAAAAGATCGAACCTTTGAGCAAGTCGAG 120  
Db 73 GATGATTCTGATGATGACGACTTCGTTCCCTAAAAAAGATCGAACCTTTGAGCAAGTCGAG 132  
QY 121 GCTATTGTCAGAACTGATGCGAAAGAAAATGCATGTCAGGCTTGTTGGGGAAGTACTAAT 180  
Db 133 GCTATTGTCAGAACTGATGCGAAAGAAAATGCATGTCAGGCTTGTTGGGGAAGTACTAAT 192  
QY 181 CTTGTAAGCTGCAATACATGCACCTTATCGGTTCCATGCTAAATGCTTAGTCCACCTCTT 240  
Db 193 CTTGTAAGCTGCAATACATGCACCTTATCGGTTCCATGCTAAATGCTTAGTCCACCTCTT 252  
QY 241 AAAGATGCTTCCGTGGAAAAATTGGAGATGCCCTGAATGTGTTAGTCCCTTTAACCGAGATA 300  
Db 253 AAAGATGCTTCCGTGGAAAAATTGGAGATGCCCTGAATGTGTTAGTCCCTTTAACCGAGATA 312  
QY 301 GATAAGATATTGGATTGTGAAATCGTCCCTACAAAATCTAGTGAACAAGGTTCTCCCGAT 360  
Db 313 GATAAGATATTGGATTGTGAAATCGTCCCTACAAAATCTAGTGAACAAGGTTCTCCCGAT 372  
QY 361 GCGGAACCGAAGCCCAATTTTGTGAAACAGTATCTCGTGAAGTGAAGGGATTATCATAC 420  
Db 373 GCGGAACCGAAGCCCAATTTTGTGAAACAGTATCTCGTGAAGTGAAGGGATTATCATAC 432  
QY 421 CTTCACTGCTCTTGGGTGCCTGAGAAAGGAGTTCCAGAAGGCTTATAAGTCAATCATCGT 480  
Db 433 CTTCACTGCTCTTGGGTGCCTGAGAAAGGAGTTCCAGAAGGCTTATAAGTCAATCATCGT 492  
QY 481 TTAATAAACACAGAGTGAAACAATTTTCACCGTCAATGGAGTCATTCAATAACAGCGAAGAT 540  
Db 493 TTAATAAACACAGAGTGAAACAATTTTCACCGTCAATGGAGTCATTCAATAACAGCGAAGAT 552  
QY 541 GATTTTGTGCCATACGTCCTGAGTGGACCACCTGTTGATCGGATTCTTGCCTGCAGAGAG 600

Db 553 GATTTTGTGCCATACGTCCTGAGTGACCACCTGTTGATCGGATTCTTGCCTGCAGAGAG 612  
QY 601 GAAGATGGGAGCTGGAATATCTTGTCAAATATAAAGAGCTATCCTATGATGAATGTTAT 660  
Db 613 GAAGATGGGAGCTGGAATATCTTGTCAAATATAAAGAGCTATCCTATGATGAATGTTAT 672  
QY 661 TGGGAGTCAGAATCAGACATCTCAACCTTCCAGAATGAAATTCAAAAGGTTCAAGGATGTA 720  
Db 673 TGGGAGTCAGAATCAGACATCTCAACCTTCCAGAATGAAATTCAAAAGGTTCAAGGATGTA 732  
QY 721 AATTCTAGAACTCGCAGAAAGTAAAGATGTTGACCATAAAAAGAAATCCCAGAGACTTTCAA 780  
Db 733 AATTCTAGAACTCGCAGAAAGTAAAGATGTTGACCATAAAAAGAAATCCCAGAGACTTTCAA 792  
QY 781 CAGTTTGATCATACTCCTGAATTCCTCAAAAGGCTTGTTACATCCATACCAGCTTGAGGGA 840  
Db 793 CAGTTTGATCATACTCCTGATTCCTCAAAAGGCTTGTTACATCCATACCAGCTTGAGGGA 852  
QY 841 CTTAATTTTTCGGGTTCTCGTGGTCAAAAACAGACGCATGTAATCCTTGTGATGAAATG 900  
Db 853 CTTAATTTTTCGGGTTCTCGTGGTCAAAAACAGACGCATGTAATCCTTGTGATGAAATG 911  
QY 901 GGACTAGGCAAGACAAATTCAAAGCATTG 928  
Db 912 GGACTAGGCAAGACAAATTCAAAGCATTG 939

RESULT 2  
AU237917  
LOCUS AU237917 RAFL16 Arabidopsis thaliana cDNA clone RAFL16-66-N20 5',  
DEFINITION mRNA sequence.  
ACCESSION AU237917  
VERSION AU237917.1 GI:19877086  
KEYWORDS EST.  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 639)  
Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,  
Akiyama,K., Enju,A., Oono,Y., Sakurai,T., Carninci,P., Kawai,J.,  
Itoh,M., Ishii,Y., Arakawa,T., Shibata,K., Shinagawa,A.,  
Muramatsu,M., Hayashizaki,Y. and Shinozaki,K.  
Large scale analysis of Arabidopsis full-length cDNA  
Unpublished (2002)  
Contact: Motoaki Seki  
Plant Functional Genomics Research Group  
RIKEN Genomic Sciences Center  
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel: 81-298-36-4359  
Fax: 81-298-36-9060  
Email: mseki@rtc.riken.go.jp  
An Arabidopsis full-length cDNA library was constructed essentially  
as reported previously (Seki et al., 1998). cDNA cleaved with BamHI  
and XhoI was ligated to modified lambda FLC-1 vector (Carninci et  
al., submitted for publication) digested with BamHI and Sali. This  
clone is in a modified pBluescript vector. Please visit our web  
site (http://www.gsc.riken.go.jp/e/plant/index\_e.html) for further  
details.

FEATURES  
Location/Qualifiers  
1. .639  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/db\_xref="taxon:3702"  
/clone="RAFL16-66-N20"  
/lab\_host="DH10B"  
/clone\_lib="RAFL16"  
/note="Site\_1: BamHI; Site\_2: Sali; dark-grown"

ORIGIN



Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnurch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257

12477932

2 (bases 1 to 4226)

Strausberg,R.

Direct Submission

Submitted (01-OCT-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)

Tissue Procurement: Dr. Jim Lin, University of Iowa

CDNA Library Preparation: M. Bento Soares, University of Iowa

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada

[info@bcgsc.bc.ca](mailto:info@bcgsc.bc.ca)

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAK Plate: 132 Row: n Column: 18

This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis

This clone has the following problem: frame shifted.

Location/Qualifiers

1. .4226

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6"

/db\_xref="taxon:10090"

/clone="IMAGE:6827288"

/tissue\_type="Brain, mouse, 13.5,14.5,16.5,17.5 dpc"

/clone\_lib="NIH\_BMAP\_FY0"

/lab\_host="DH10B"

/note="Vector: pyX-ASC"

FEATURES

source

ORIGIN

Query Match 9.9%; Score 415.4; DB 11; Length 4226;

Best Local Similarity 52.8%; Pred. No. 7.6e-83;

Matches 1261; Conservative 0; Mismatches 1026; Indels 100; Gaps 13;

QY 818 TACATCCATACGAGCTTGAGGGACTTAATTTTTCGGGTTCTCGTGGTCAAAACAGACGC 877

Db 114 TGCATCCCTACCAATGGAAGGCTTAACTGGCTCCGCTTCTCTCGGCTCAGGGAACCG 173

QY 878 ATGTAATCCTTGCTGATGAAATGGGACTAGGCAAGACAATTCAAAGCATTGCCCTTTTAG 937

174 ATACTATCTTGGCTGATGAGATGGGCCTTGGGAAAACTGTGCAGACAGCAGTCTTCTCT 233

QY 938 CTTCACTTTTGGAGAGAACCTCAATT-----CCGCATTTGGTAATTGCTCCTCTATCGA 991

Db 234 ACTCCCTCTATAAGGAGGCTCACTCCAAAGGCCCTTCTTAGTGAGCGCTCCTCTGTCCA 293

QY 992 CTCTGCGTAACCTGGGAGAGAGAGTTTGGCCACATGGSCCCACACAGATGAACGTGGTATGT 1051

Db 294 CCATCATCAACTGGGAGAGAGAGTTTGAGATGTGGCTCCTGATATGTATGTGTTAACCT 353

QY 1052 ATTTTGGCACTGGCGAAGCTCGAGCAGTTTATCAGAGAACATGAGTTTATCTATCGAAAG 1111

Db 354 ATGTGGGTGACAAGGACAGCCGTGCCATCATCCGGGAAAATGAGTTCTCCTTTGAAGACA 413

QY 1112 ATCAAAAAAAGATCAAGAAAAAGAAATCTGGACAATAAGTAGCGAAAGCAAGCAAAAAA 1171

Db 414 ATGCCATTTCGTGGTGGCAAGAGGCCTC-----GAGGATGAAGAAAGAGCAT 461

QY 1172 GAATCAAGTTTGTATGTCCTCTCACATCGTATGAGATGATCAACCTAGATTTCAGCAGTTC 1231

Db 462 CGTGAAATTCATGTTCTGCTGACATCCTATGAGTTGATCACCATTTGATATGATGCCATCT 521

QY 1232 TAAACCAATTAAGTGGGAGTGCATGATTGTTGATGAAGGTCATCGACTGAAAAATAAGG 1291

Db 522 TGGGTTCTATTGACTGGGCTGCCTCATTTGTGGTAGAGCCCATCGGCTGAAGAACCAACC 581

QY 1292 ATTCAAAAGCTGTTCTCTTATTGACACAGTATTCAAGTAACCAACCGTATTCTTCTGACAG 1351

Db 582 AGTCTAAGTTCTTTTCGAGTTTGAATGGTTACTCACTCCAGCACAAAGCTGTTGCTAACTG 641

QY 1352 GAAACACACTTCAGAACAACTTGGATGAACCTTTTTCATGCTCATGCAATTTTCTTGTATGCGG 1411

Db 642 GAACTCCGTTACAGAACAACTAGAGGAACTGTTTCACTGCTCAACTTCTCACCCCTG 701

QY 1412 GGAAGTTTGAAGTTTGGAGGAGTTCCAGGAGGAGTTCCAGGAGGAGTTCAAAGATATTAAATCAAGAGGAGC 1471

Db 702 AGAGGTTCCACAACCTTAGAAGGCTTCTTGGAGGAGTTTGCAGACATTTGCCAAAGAGGACC 761

QY 1472 AGATCTCAAGGTTGCACAAAATGTTGGCTCCACATTTTGTCTCAGAAAGGGTAAAAAAGACG 1531

Db 762 AGATTAAAAAACTCCACGCAGATGCTGGGCTCATATGTTGCGGCGGCTCAAAGCTGACG 821

QY 1532 TAATGAAAGACATGCCCCCAAAAAGGAGCTCATTTTTCGGTGTGATCTGAGCAGTCTGC 1591

Db 822 TCTTCAAGAATATGCCATCCAAGACAGAGCTGATTGTCGGTGGAGTTGAGTCCATATGC 881

QY 1592 AGAAAGAATATTACAAAGCTATTTTACCCGTAATTTATCAAGTATTGA--CAAAAAAGG 1648

Db 882 AGAAGAAGTACTACAAGTATATTCTCAGCGGGAATTTTCAAGACACTTAATGCTCGAGGTG 941

QY 1649 GAGTGTCTCAAATTTCCCTTAATAACATTATGATGGAATTACGAAAAGTATGCTGCCATC 1708

Db 942 GTGGCAACCAGGTTTCTCTGCTGAAACGTGGTATGATGATCTTAAGAAAATGCTGCAACCACC 1001

QY 1709 CTTA-----TATGCTAGAGGTTGTTGAGCCAGTTTATTCACGACGCAAAATGAAGCTT 1759

Db 1002 CTTATCTCTTCCCTGTGGCAGCAATGGAAGCCCCCTAAGATGCCTAAATGGTATGATGATG 1061

QY 1760 TCAACAACACTTTTGGAGCTTCTGTGAAAAGCTGCAACTTCTAGATAAAAATGATGGTCAAAC 1819

Db 1062 GCAGTCCCTTAATCAGAGCATCTGGGAAGTTGTTGCTGCTGCAGAAGATGCTTAAGAACC 1121

QY 1820 TGAAGAGCAAGGACACAGAGTCTCTAATATACACACAGTTTTCAGCATATGCTGGACTTAC 1879

Db 1122 TGAAGAGGGAGGGCATCGTGTGCTCATCTTCTCCAGATGACCAAGATGTTGGACTTGC 1181

QY 1880 TTGAAGACTACTGTACCCATAAGAAATGGCAGTACGAGCGAATTGATGGAAAGGTTGGCG 1939

Db 1182 TAGAGGATTTTTAGAACATGAAGGTTATAAATATGAACGTTATGATGGTGAATCACTG 1241

QY 1940 GAGCTGAGCGGCAAAATACGATFAGATCGGTTCAATGCCAAAAAATCTTAACAAGTTTGT 1999



Db 1242 GGAACATGCGTCAGGAGGCCATTGATCGCTTCAATGCACCGGFRGCTCAACAGTTCTGCT 1301

Qy 2000 TTTTGTCTCCACAAGAGCTGGTGGCTTAGGAATAAAATCTTGCAAGGCTGATACAGTAA 2059

Db 1302 TCTTGTCTTCCACTCGAGCTGGGGCCTTGGGATCAATCTGCCACTCGGCACACAGTAA 1361

Qy 2060 TCATTTATGACAGTGACTGGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC 2119

Db 1362 TTATATATGACTCTGACTGGAACCCCATATGA-CATCAGGCTTTAGCAGAGCCCAACC 1420

Qy 2120 GACTTGGCCAAACAATAAGGTGATGATTTATAGGCTCATAAACCGAGGCCACCATTTGAAG 2179

Db 1421 GTATTGGGCAANAATAAGAAGGTGATGATCTACCGGTTTGTGACCCGTGCATCAGTGGAGG 1480

Qy 2180 AAAGGATGATGCAATTGACTAAAGAGAAAATGGTTCTAGAGCATCTTTGTTGTTGG----- 2234

Db 1481 AGCGCATACGCAGGTGGCAAAGAAAGAAAGATGATGCTGACACATTTAGTGGTACGGCCTG 1540

Qy 2235 ---GAAACTCAAAACACAAAACATTAATCAGGAAGAGTTAGATGACATCATCAGGTATG 2290

Db 1541 GCCTGGGCTCCAAGACAGGCTCAATGTCTAAACAGGAGCTTGATGATATCCTCAAATTTG 1600

Qy 2291 GATCAAGGAGCTTTTGTAGTGAAG-----ATGATGAAGCAG 2329

Db 1601 GCACTGAGAGCTATTCAAGGATGAAGCCAGGATGGAGGAGGAGACAAAGGAGGGAG 1660

Qy 2330 GAAAGCTCGAAAAAATTCATTATGATGATGCGGCTATAGACAAATTTGCTGATCGTGATC 2389

Db 1661 AAGACAGCAGTGTATCCACTATGACGATAAGGCCATTGAACGACTGCTGGATCGAAACC 1720

Qy 2390 TCGTGGAGGCAGAGGAAGTCTCAGTGGATGATGAAGAGGAGATGGATTCTTAAAGGCTT 2449

Db 1721 AGGATGAGACTGAAGACACACAGAATTG-----CAGGGCATGAATGAATATTTGAGCTCAT 1774

Qy 2450 TCAAGGTGGCTAAATTTTGAATATATAGATGAAAAATGAGGCAGCAGCATTAGAGGCACAGA 2509

Db 1775 TCAAAAGTGGCTCAGTATGTGGTACGGGAAGAAGAGATGG-----GGGAGGAAGAGG 1825

Qy 2510 GAGTCGCTGCTGAAAGCAAATCTTCAGCAGGCAATTTCTGATAGAGCAAGTTATTGGGAAG 2569

Db 1826 AGGTAGAACGGGAATCATAAACAGGAAGAAAGTGTGGATCTTGAC---TACTGGGAGA 1882

Qy 2570 AGTTGTTAAAGATAAATTTGAGCTGCACACAGGCTGAGGAGCTTAATGCTCTTGGAAAAA 2629

Db 1883 AATTGCTCGGCACCATTTATGAGCAGCAGCAAGAAGATCTAGCCCCGAAATCTGGGCAAG 1942

Qy 2630 GGAAGACAAGTCGCAAGCAGTTGGTATCCATTGAAG---AAGATGATCTTGTGTTGGTGG 2686

Db 1943 GAAAAAGAAATCCGTAACAGGTCAACTACAATGATGGCTCAGAGGAGCCGAGATTGGC 2002

Qy 2687 AAGATGTGAGCTCTGATGGAGATGAAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAG 2746

Db 2003 AGGACGACCAGTCCGACAACCAAGTCCGATTACTCAGTGGCCTCAGAGGAAGGTGATGAAG 2062

Qy 2747 CAGGACAAGGAGTTCAGACGGGTCGACGGCCGTACAGAAGAAAGGG-----TCGCG 2797

Db 2063 ACTTTGACGAACGGTCAGAAGCTCCCCGAGGCCCAGTCCGAAGGGCCTGCGGAATGATA 2122

Qy 2798 ATAATTTGGAACCAACTCCGTTGATGGAAGGTGAGGGGAGATCTTTTCAGAGTACTGGTT 2857

Db 2123 AAGATAAGCCATTACCTCCTCTGTTGGCCCGTGTGGTGGGAATATTGAAGTACTTGGTT 2182

Qy 2858 TCAACACAGAGTCAAAGGGCCATTTTGTACAGACTTTGTATGAGGTATGG-----AG 2908

Db 2183 TTAATGCTGTCAGCGAAAAAGCATTTCTTAATGCAATTTATGCGATATGGCATGCCACCTC 2242

Qy 2909 CTGGCAATTTTGAATTGGAAGGAGTTGTTCTCGCTTAAAGCAGAAAGACCTTTTGAAGAAA 2968

Db 2243 AGGATGCTTTTACCACACTCAGTGGCTGTGAGAGATCTTTCGAGSCAAGTCAGAGAAGAAT 2302

Qy 2969 TAAATGAATATGGAATACTCTTCTGAAGCACATTGCTGAAGAAATAGACGAGAATTCTC 3028

Db 2303 TTAAGGCTTATGTGTCACTCTTCATGCGACATTTGTGTGAGCCTGGGGCAGATGGGCGTG 2362

Qy 3029 CAACCTTTTTCAGATGGTGTGCCCCAAGGAAGGACTTAGAATAGAAGATGTTCTAGTCAGAA 3088

Db 2363 AGACCTTTTGTGATGGCTGCCACGAGAAGGCCCTGTCTCGTCAACAGTTTCTCACTAGGA 2422

Qy 3089 TTGCTCTTCTGATACTAGTTCAGGAGAAAGGTGAAATTTGTAGAAGAT 3135

Db 2423 TCGGTGTCATGTCTCTGATTCGAAAGAAGGTTTCAGGAGTTTGAACAT 2469

RESULT 5  
AK034549  
LOCUS  
DEFINITION  
AK034549 3137 bp mRNA linear HTC 18-SEP-2003  
Mus musculus 12 days embryo embryonic body between diaphragm region  
and neck cDNA, RIKEN full-length enriched library, clone:9430004K15  
product:CHROMODOMAIN HELICASE-DNA-BINDING PROTEIN 4 (CHD-4) (MI-2  
AUTOANTIGEN 218 KDA PROTEIN) (MI2-BETA) homolog [Homo sapiens],  
full insert sequence.

ACCESSION  
AK034549  
VERSION  
AK034549.1 GI:26330020  
KEYWORDS  
HTC; CAP trapper.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus

REFERENCE  
1  
Carninci,P. and Hayashizaki,Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
JOURNAL  
99279253  
MEDLINE  
10349636  
PUBMED  
REFERENCE  
2

AUTHORS  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,  
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
JOURNAL  
20499374  
MEDLINE  
11042159  
PUBMED  
REFERENCE  
3

AUTHORS  
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,  
Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,  
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,  
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,  
Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,  
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,  
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
JOURNAL  
20530913  
MEDLINE  
11076861  
PUBMED  
REFERENCE  
4

AUTHORS  
The RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 585-590 (2001)  
JOURNAL  
20530913  
MEDLINE  
11076861  
PUBMED  
REFERENCE  
5

AUTHORS  
The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
JOURNAL  
20530913  
MEDLINE  
11076861  
PUBMED  
REFERENCE  
6

AUTHORS  
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,  
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,  
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,  
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,  
Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,  
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,  
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,  
Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,  
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,



[illegible]

RESULT 6	
BE204297	
LOCUS	610 bp mRNA linear EST 05-SEP-2000
DEFINITION	EST396973 KV0 Medicago truncatula cDNA clone pKV0-15G16, mRNA sequence.
ACCESSION	BE204297
VERSION	BE204297.1 GI:8747582
KEYWORDS	EST.
SOURCE	Medicago truncatula (barrel medic)
ORGANISM	Medicago truncatula Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago. 1 (bases 1 to 610) VandenBosch,K., Endre,G., Hur,J., Moore,J., Beremand,P., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.
REFERENCE	
AUTHORS	

```

FEATURES             source
Location/Qualifiers
1. .610
  /organism="Medicago truncatula"
  /mol_type="mRNA"
  /cultivar="genotype Al7"
  /db_xref="taxon:3880"
  /clone="pKV0-15G16"
  /tissue_type="Seedling roots"
  /dev_stage="Immediately prior to inoculation with
Sinorhizobium meliloti (0 hour)"
  /lab_host="E.coli strain XL0LR"
  /clone_lib="KV0"
  /note="Vector: pBluescript SK -; Site_1: EcoRI; Site_2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the Unizap XR vector from

```

Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XL0LR cells."

Query Match	9.5%;	Score 398.8;	DB 10;	Length 610;
Best Local Similarity	78.4%;	Pred. No. 3.1e-79;		
Matches 478;	Conservative 0;	Mismatches 132;	Indels 0;	Gaps 0;
QY	1570	CGTGTGATCTGAGCAGTCTGCAGAAAGAATATTACAAAAGCTATTTTACC CGTAAATTAT	1629	
Db	1	CGAGTGGATTGAGCAGCAAAACAGAAAGAATATTATAAGGCAATTTTGACCCGTAATTAT	60	
QY	1630	CAAGTATTGACAAAAAAGGAGGTGCTCAAATTTCCCTTAATAACATTATGATGGAATTA	1689	
Db	61	CAGATATTAAACCGCCGTTGGTGTGCACAGATTTCTTTATCAACGTTGTTATGGAAC TG	120	
QY	1690	CGAAAAGTATGCTGCCATCCTTATATGCTAGAGGGTGTGTGAGCCAGTTATTCACGACGCA	1749	
Db	121	CGTAAGCTCTGTTGTTCATGCTTACATGTTAGAAGGAGTTGAGCCAGATATTGACGATCCA	180	
QY	1750	AATGAAGCTTTCAAACAACCTTTTGGAGTCTTGTGGAAAGCTGCAACCTTCAGATAAAATG	1809	
Db	181	AAAGAAGCATTCAGCAATTGCTGGAATCATCAGGGAAGTTGCACCTTGTCTGACAAGATG	240	
QY	1810	ATGGTCAAACTGAAAGAGCAAGGACACAGAGTCCTAATATACACACAGTTTCAGCATATG	1869	
Db	241	ATGGTGAACTTAAAGAACCAAGGACATAGAGTTCTCATATATCTCCAGTTTCAGCACATG	300	
QY	1870	CTGGACTTACTTGAAGACTACTGTACCCATAAGAAATGGCAGTACGAGCGAATTGATGGA	1929	
Db	301	CTTGATTGCTTGAAGATTACTGTCTTTACAAGAAATGGCATTATGAAGGATAGATGGC	360	
QY	1930	AAGTTGGCGGAGCTGAGCGGCAATACGCATAGATCGTTTCAATGCCAAAAAATTTCTAAC	1989	
Db	361	AAGTTGTTGGGGCTGAAAGACAAATACGGATAGATCGTTTAAATGCCAAAAAATTTCTTCA	420	
QY	1990	AAGTTTGTGTTTTTGTCTCTCCACAAGAGCTGGTGGCTTAGGAATAAATCTTGCAACGGCT	2049	
Db	421	AGATTTTGTCTTCTTCTTCTACAAGAGCCGGGGTTTGGGAATAAACCTTGCAACTGCT	480	
QY	2050	GATACAGTAATCATTTATGACAGTGACTGGAATCCTCATGCTGATCTTCAAGCAATGGCT	2109	
Db	481	GACACAGTTGTTATTTATGACAGTGATTGGAATCCTCATGCTGACCTACAAGCAATGGCT	540	
QY	2110	AGAGCTCATCGACTTGGCCAAAACAATAAGGTGATGATTTATAGGCTCATAAACCGAGGC	2169	
Db	541	AGAGCTCATCGACTTGGACAAACTAACAAAGGTGTTGATTTATAGGCTTATAACACGGGGA	600	
QY	2170	ACCATTGAAG	2179	
Db	601	ACGATTGAAG	610	

RESULT 7	CB635984	LOCUS	CB635984	745 bp	mRNA	linear	EST 08-APR-2003
DEFINITION	OSIIEb16P12.f OSIIEb <i>Oryza sativa</i> (indica cultivar-group) cDNA clone OSIIEb16P12 5', mRNA sequence.						
ACCESSION	CB635984						
VERSION	CB635984.1	GI:29630975					
KEYWORDS	EST.						
SOURCE	<i>Oryza sativa</i> (indica cultivar-group)						
ORGANISM	<i>Oryza sativa</i> (indica cultivar-group)						
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; <i>Oryza</i> .						
REFERENCE	1 (bases 1 to 745)						
AUTHORS	Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E., Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.						
TITLE	Large-scale identification of ESTs involved in the interaction between rice and Magnaporthe oryzae						



JOURNAL COMMENT	Unpublished (2003) Contact: Rod Wing Arizona Genomics Institute University of Arizona Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ 85721-0088, USA Tel: 520 626 3967 Fax: 520 621 9288 Email: http://genome.arizona.edu PCR Primers FORWARD: gta aaa cga cgg cca gtcg BACKWARD: gga aac agc tat gac cat g Plate: 16 row: P column: 12 Seq primer: gta aaa cga cgg cca gtcg.
FEATURES	Location/Qualifiers
source	1. .745 /organism="Oryza sativa (indica cultivar-group)" /mol_type="mRNA" /cultivar="IR36" /db_xref="taxon:39946" /clone="OSIIEb16P12" /tissue_type="Leaf" /dev_stage="3 week" /lab_host="DH10B" /clone_lib="OSIIEb" /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2: XhoI; 24 hrs after inoculation with Rice Blast (PO6-6-3)"
ORIGIN	
Query Match	9.3%; Score 387.6; DB 14; Length 745;
Best Local Similarity	72.9%; Pred. No. 1.1e-76;
Matches 541; Conservative	0; Mismatches 194; Indels 7; Gaps 3;
QY	1661 TTTCCTTAATAACATTATGATGGAATTACGAAAAGTA-TGCTGCCATCCTTATATGCTA 1719 
Db	4 TTTCACCTTATTCATGAAGAATGGAACCTACTCAAACCTATTGTTGCCATGCATTCATGACA 63 
QY	1720 GAGGGTGTGACCAGTTATTTCACGACGCCAAATGAAGCTTTTCAAACAACCTTTTGGAGTCT 1779 
Db	64 GATGAACCCGA---AGAGCCTGCCAATTTCAGAAGAAGCTTTTAAGGAGGCTTTTAGAATCT 120 
QY	1780 TGTGAAAGCTGCAACTTCTAGATAAAATGATGGTCAAACTGAAAGACCAAGCACACAGA 1839 
Db	121 TCTGAAAAAATGGAGCTGCTTGACAAGATGATGGTGAAACTGAAAGACGAGGTCACAGG 180 
QY	1840 GTCCTAATATACACAGATTTTCAGCATAATGCTGGACTTACTTGAAGACTACTGTACCCAT 1899 
Db	181 GTTCTTATTTATTCACAGTTCACAGTCACATGTTGGACTTACTTGAGGATTTATTTAAGCTAC 240 
QY	1900 AAGAAATGGCAGTACGACGGAATTGATGGAAGGTTGGCGGAGCTGAGCGGCAAAATACGC 1959 
Db	241 CGGAAATGGAGTTATGAGCGGTATTGATGGAAAAATAGGTGGTGCTGAGAGGCAGATACGA 300 
QY	1960 ATAGATCGGTTCAATGCCAAAAATTCTAACAGTTTGTGTTTTTGTCTCTCCACAAGAGCT 2019 
Db	301 ATTGATCGCTTCAATGCTAAAAATTTCTACTAGGTTTGTCTTCTTCTTCAACCAGAGCT 360 
QY	2020 GGTGGCTTAGGAATAAATCTTGCAACGGCTGATACAGTAATCATTTATGACAGTGACTGG 2079 
Db	361 GGTGGTCTTGGTATAAATCTGGCAACTGCCGATACTGTAATTATTTATGACAGTGATTGG 420 
QY	2080 AATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATCGACTTGGCCAAACAATAAG 2139 
Db	421 AATCCTCATGCTGATTTACAAGCTATGGCGAGAGCCCATCGCTTAGGGCAGACAAGCAAG 480 
QY	2140 GTGATGATTTATAGGCTCATAAACCGAGGCACCATTTGAAGAAAGGATGATGCAATTGACT 2199 
Db	481 GTTATGATATATAGGCTTGTAGTCGTGTTACAACTGCCGATACTGTAATTATTTATGACAGTGATTGG 540 
QY	2200 AAAAAAGAAAATGGTTCATAGAGCATCTTGTGTTGGGAAACTC---AAAAACAAAAACATT 2256 
Db	541 AAGAAAAAAATGGTATTGGAGCACTTAGTTGTTGGGCGTCTCACGAAAGGCACCTAATATT 600 
QY	2257 AATCAGGAAGAGTTAGATGACATCATCAGGTATGGATCAAAAGGAGCTTTTTTGCTAGTGAA 2316 
Db	601 GTCCAGGAGGAGTTGGATGATATATTTCGGCATGGCTCAAAGGAACTTTTTTGACGATGAA 660 
QY	2317 GATGATGAAGCAGGAAAGTCTGGAAAAATTCATTATGATGATCGGGCTATAGACAAATTG 2376 
Db	661 AATGATGAAGCCGGAAAATCTTGCCAAATCCATTATGATGATGCTGCGATTGATAAATTA 720 
QY	2377 CTTGATCGTGATCTCGTGGAGG 2398 
Db	721 TTGGACCGTGACCAAGCTGACG 742 
RESULT 8	
AI998628/c	
LOCUS	AI998628 563 bp mRNA linear EST 08-SEP-1999
DEFINITION	701546371 A. thaliana, Columbia Col-0, rosette-2 Arabidopsis thaliana cDNA clone 701546371, mRNA sequence.
ACCESSION	AI998628
VERSION	AI998628.1 GI:5845533
KEYWORDS	EST.
SOURCE	Arabidopsis thaliana (thale cress)
ORGANISM	Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE	1 (bases 1 to 563) Chen,J., Momiyama,M., Chan,E., Mooney,M., Carroon,B., Gilliland,D., Wang,X., Hillman,J., Guegler,K., Kim,C., Doyle,M., Brzoska,P., Gorgone,G., Burns,D., Griffin,J., Mouanoutoua,M., Nguyen,D., Tan,R., Rose,M., Warren,B., Ton,B., Kastury,K., Borillo,C., Carpio,T., Policky,J., Suzuki,G., Argentine,C., Shah,S., Nobriga,A., Murry,L., Turner,C., Krikorian,S., Elder,L. and Hanson,D.
AUTHORS	
TITLE	Arabidopsis thaliana Gene Expression MicroArray
JOURNAL	Unpublished (1999)
COMMENT	Contact: David Smoller, Ph.D. Genome Systems, Inc., a wholly owned subsidiary of Incyte Pharmaceuticals, Inc. 4633 World Parkway Circle, St. Louis, MO 63134, USA Tel: 877-577-2733 Fax: 314-427-3324 Email: service@genomesystems.com.
FEATURES	Location/Qualifiers
source	1. .563 /organism="Arabidopsis thaliana" /mol_type="mRNA" /cultivar="Columbia Col-0" /db_xref="taxon:3702" /clone="701546371" /tissue_type="rosette" /dev_stage="4 - 7 weeks" /clone_lib="A. thaliana, Columbia Col-0, rosette-2" /note="Vector: pSPORT; Site 1: NotI; Site 2: SalI; cDNA library was derived from untreated rosette tissue from Arabidopsis thaliana, Columbia Col-0, at 4 - 7 weeks. Plants were grown in 1:1:1 peat moss/vermiculite/perlite soil at 22 deg. C +/- 3 deg. C under constant light, and watered with fertilizer. cDNA synthesis was initiated using a NotI-oligo(dT) primer. Double-stranded cDNA was blunted, ligated to SalI adaptors, digested with NotI, size-selected, and cloned into the NotI and SalI sites of the pSPORT vector."
ORIGIN	
Query Match	9.2%; Score 385; DB 9; Length 563;
Best Local Similarity	100.0%; Pred. No. 4.1e-76;
Matches 385; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	3793 CAGATGTGCAAACTTCTTGATGAGAACGCTCGGGAATCAGTCAAGCATATGTAAACAAC 3852 
Db	563 CAGATGTGCAAACTTCTTGATGAGAACGCTCGGGAATCAGTCAAGCATATGTAAACAAC 504 

QY 3853 CAACCACCGAGTACCAAGGTGAATGAGAGCTTCGTCGACTCAATCTATCAATGGTAAC 3912  
|||||  
Db 503 CAACCACCGAGTACCAAGGTGAATGAGAGCTTCGTCGACTCAATCTATCAATGGTAAC 444  
  
QY 3913 ATTAACACAAATCCTTTTCGATTACATCTGATCAATCCAAGTCACATGAAGACGACACCAAG 3972  
|||||  
Db 443 ATTAACACAAATCCTTTTCGATTACATCTGATCAATCCAAGTCACATGAAGACGACACCAAG 384  
  
QY 3973 CCAGACCTAAACAATGTTGAGATGAAGACACGGCCGAAGAACAACAAACCGTTAAGAGGT 4032  
|||||  
Db 383 CCAGACCTAAACAATGTTGAGATGAAGACACGGCCGAAGAACAACAAACCGTTAAGAGGT 324  
  
QY 4033 GCGTCGTCGATCTGAATGTGGTGGAGGAGGAGGAACAATGCTGAAGCTAGTGGAGT 4092  
|||||  
Db 323 GCGTCGTCGATCTGAATGTGGTGGAGGAGGAGGAACAATGCTGAAGCTAGTGGAGT 264  
  
QY 4093 GTTGATGTAAAAATGGAAGAAGCCAAAGAAGAGAGCCAAAGAACATGGTCGTTGAT 4152  
|||||  
Db 263 GTTGATGTAAAAATGGAAGAAGCCAAAGAAGAGAGCCAAAGAACATGGTCGTTGAT 204  
  
QY 4153 TGACTCACTGGTAAATCAAGATTC 4177  
|||||  
Db 203 TGACTCACTGGTAAATCAAGATTC 179

RESULT 9  
BE204298  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
Medicago.  
1 (bases 1 to 597)  
VandenBosch,K., Endre,G., Hur,J., Moore,J., Beremand,P., Ellis,L.,  
Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and  
Fraser,C.M.  
ESTs from uninoculated seedling roots of Medicago truncatula  
Unpublished (1999)  
Contact: VandenBosch K  
Department of Plant Biology  
University of Minnesota  
220 Biosci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA  
Tel: 612 624 2755  
Fax: 612 625 1738  
Email: kvandenb@cbs.umn.edu  
Texas A&M University name:T264070e  
TIGR sequence name:MTGAR45TK  
More information is available at.  
http://chrysie.tamu.edu/medicago  
Seq primer: SKmod (CTA GAA CTA gtg gat CC).  
Location/Qualifiers  
1. .597  
/organism="Medicago truncatula"  
/mol\_type="mRNA"  
/cultivar="genotype A17"  
/db\_xref="taxon:3880"  
/clone="pKV0-15G18"  
/tissue type="Seedling roots"  
/dev stage="Immediately prior to inoculation with  
Sinochizobium meliloti (0 hour)"  
/lab\_host="E.coli strain XL0LR"  
/clone lib="KV0"

FEATURES  
source  
1 (bases 1 to 589)  
Watson,B.S., Shin,H.-S., Lopez-Meyer,M., Scott,A.D., Harris,A.R.,  
Gonzales,R.A., Bell,C.J., Inman,J.T., Waugh,M.E., Sullivan,J.P.,  
May,G.D. and Paiva,N.L.  
Expressed Sequence Tags from the Samuel Roberts Noble Foundation  
Medicago truncatula Phoma-infected library  
Unpublished (2002)  
Contact: Paiva NL  
Plant Biology Division

ORIGIN

Query Match 8.9%; Score 373; DB 10; Length 597;  
Best Local Similarity 77.9%; Pred. No. 2.1e-73;  
Matches 462; Conservative 0; Mismatches 130; Indels 1; Gaps 1;  
  
QY 1570 CGTGTGATCTGAGCAGTCTGCAGAAAAGAAATATTACAAAGCTATTTTACCCTGAATTAT 1629  
|||||  
Db 2 CGAGTGGATTTCGAGCAGCAAAACAGAAAAGAAATATTATAAGGCAATTTTGACCCGTAATTAT 61  
  
QY 1630 CAAGTATTGACAAAAAAGGAGGTGCTCA-AAATTCCTTAATAACATTATGATGSAATT 1688  
|||||  
Db 62 CAGATATTAAACCCCGCGTGGTGGTGCACAGTATTTCTCTTATCAACGTTGTTATGGAAC 121  
  
QY 1689 ACGAAAAAGTATGCTGCCATCCTTATATGCTAGAGGGTGTGAGCCAGTTATTTCACGACGC 1748  
|||||  
Db 122 GCGTAAGCTCTGTTGTCATGCTTACATGTTAGAAGGAGTTGAGCCAGATATTGACGATCC 181  
  
QY 1749 AAATGAAGCTTTCAAACAACACTTTTGGAGTCTTGTGGAAAGCTGCAACTTCTAGATAAAAT 1808  
|||||  
Db 182 AAAAGAAGCATTTCAAGCAATTGCTGGAATCATCAGGGAAGTTGCACCTTGTGGACAAGAT 241  
  
QY 1809 GATGGTCAAACTGAAAGAGCAAGGACACAGAGTCCTTAATATACACAGTTTCAGCATAT 1868  
|||||  
Db 242 GATGGTGAAGCTTAAAGAAACAAGGACATAGAGTTCTCATATACTCCAGTTTCAGCACAT 301  
  
QY 1869 GCTGGACTTACTTGAAGACTACTGTACCCATAAGAAATGGCAGTACGAGCGAATTGATGG 1928  
|||||  
Db 302 GCTTGATTGCTTGAAGATTACTGCTCTTACAAGAAATGGCATATTGAAAGGATAGATGG 361  
  
QY 1929 AAAGGTTGGCGGAGCTGAGCGGCAAAATACGCATAGATCGGTTCAATGCCAAAAATTCCTAA 1988  
|||||  
Db 362 CAAGGTTGGTGGGCTGAAAGACAAATACGGATAGATCGTTTTTAATGCCAAAAATTCCTC 421  
  
QY 1989 CAAGTTTGTGTTTTGCTCTCCACAAGAGCTGGTGGCTTAGGAATAAATCTTGCAACGGC 2048  
|||||  
Db 422 AAGATTTTGCTTTCTCTTCTACAAGAGCCGGGGTTTGGGAATAAACCTTGCAACTGC 481  
  
QY 2049 TCATACAGTAATCATTTATGACAGTGACTGGAATCCTCATGCTGATCTTCAAGCAATGGC 2108  
|||||  
Db 482 TGACACAGTTGTTATTTATGACAGGGATTGGAATCCTCATGCTGACCTACAAGCAATGGC 541  
  
QY 2109 TAGAGCTCATCGACTTGGCCAAAACAAATAAAGGTGATGATTTATAGGCTCATAA 2161  
|||||  
Db 542 TAGAGCTCATCGACTTGGACAAACTAACAAAGGTGTTGATTATTAAGCTTATAA 594

RESULT 10  
BQ139596  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Medicago truncatula (barrel medic)  
Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
Medicago.  
1 (bases 1 to 589)  
Watson,B.S., Shin,H.-S., Lopez-Meyer,M., Scott,A.D., Harris,A.R.,  
Gonzales,R.A., Bell,C.J., Inman,J.T., Waugh,M.E., Sullivan,J.P.,  
May,G.D. and Paiva,N.L.  
Expressed Sequence Tags from the Samuel Roberts Noble Foundation  
Medicago truncatula Phoma-infected library  
Unpublished (2002)  
Contact: Paiva NL  
Plant Biology Division

The Samuel Roberts Noble Foundation  
2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
Tel: 580 221 7317  
Fax: 580 221 7380  
Email: nlpaiva@noble.org  
Insert Length: 589 Std Error: 0.00  
Plate: 021 row: H column: 12  
Seq primer: TCACACAGGAAACAGCTATGAC.  
Location/Qualifiers

FEATURES  
source

1. .589  
/organism="Medicago truncatula"  
/mol\_type="mRNA"  
/db\_xref="taxon:3880"  
/clone="NF021H12PH"  
/tissue\_type="leaf"  
/dev\_stage="Pathogen-induced, young trifoliolate"  
/clone\_lib="Phoma-infected"  
/note="Vector: pBluescript SK(-); Young trifoliolate leaves of Medicago truncatula were excised and dip-inoculated in a spore suspension of Phoma medicaginis, and incubated in humid dishes. Pools of leaves were harvested at 0, 15, and 30 minutes and 1, 2, 3, 6, 14, 24, 48, 72, and 96, hours, and used to prepare total RNA. cDNA was prepared from polyA+ enriched, pooled samples of equivalent amounts of total RNA from each sample. The cDNA was directionally ligated into the Uni-Zap XR vector (Stratagene) and packaged using the Gigapack III Gold packaging extracts. Phagemids containing cDNA inserts were in vivo excised from the recombinant Uni-ZAP XR vector using ExAssist helper phage and the E. coli strain XL1-Blue MRF' (Stratagene). Excised plasmids were plated using SOLR cells."

ORIGIN

Query Match 8.5%; Score 354.6; DB 13; Length 589;  
Best Local Similarity 77.6%; Pred. No. 3.1e-69;  
Matches 458; Conservative 0; Mismatches 120; Indels 12; Gaps 2;  
  
QY 919 CAAAGCATTCGCCCTTTTAGCTTTCACCTTTTGGAGGAGAACCTC-----ATTCCGCATTTG 972  
Db 1 CAAAGTATTGCTTTCTTAGCATCCCTTTTGAAGAGGGTGCTCTGCACATCCACATCTG 60  
  
QY 973 GTAATTGCTCCTCTATCGACTCTGCGTAACTGGGAGAGAGAGTTTGCCACATGGGCCCA 1032  
Db 61 GTGGTTGCTCCACTTTCACACCTGCGAAACTGGGAACGTGAATTGCAACATGGGCCCT 120  
  
QY 1033 CAGATGAACGTGGTTATGTATTTTGGCACTGCGGAAGCTCGAGCAGTGTATCAGAGAACAT 1092  
Db 121 CAAATGAATGTTATTATGTATGTTGGATCTGCCCAAGCTCGTAGTGTATCAGAGAATAT 180  
  
QY 1093 GAGTTTTACTTATCGAAAGATCAAAAAGATCAAGAAAAGAAATCTGGACAAATAAGT 1152  
Db 181 GAATTTTACTTTCCCAAGAAACTGAAGAGAAACAAGAAAAGAAATCT-----TTAGTT 234  
  
QY 1153 AGCGAAAGCAAGCAAAAAAGAAATCAAGTTTGATGTCTCTCCTCACATCGTATGAGATGATC 1212  
Db 235 AGTGAAGTAAGCATGACAGGATTAAAGTTTGATGTCTTTTGACATCATATGAGATGATC 294  
  
QY 1213 AACCTAGATTACGAGTTCTAAACCCTAATTAAGTGGGAGTGCATGATTGTTGATGAAGGT 1272  
Db 295 AACTTAGACACAAACATCATTTAAACCTATAAAATGGGAGTGCATGATAGTTGATGAAGGT 354  
  
QY 1273 CATCGACTGAAAAATAAGGATTCAAAGCTGTCTCTTCATTGACACAGTATTCAAGTAAC 1332  
Db 355 CACCGCCTCAAAAATAAGGATTCAAAATATTCTTCATTGAAGCAATATTCTACCAGA 414  
  
QY 1333 CACCGTATTCTTGACAGGAACACCACCTTCAGAACAACTTGGATGAACCTTTTCATGCTC 1392  
Db 415 CATCGTGTCTCTTGACTGGAACTCCTCTTCAGAACAACTTGGATGAACCTTTTATGCTT 474  
  
QY 1393 ATGCATTTTCTTGATCGGGGAAGTTTGGAAGTTTGGAGGAGTTCAGGAGGAGTTTCAAA 1452  
Db 475 ATGCATTTCTTGATCGGGGAAGTTTGCAAGTTTANAGGAATTCCAAGAAGAATTCAAG 534

QY 1453 GATATTAATCAAGAGGAGCAGATCTCAAGTTTGACAAAATGTTGGCTCC 1502  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||  
Db 535 GATATCAATCAAGAGGAACAGATTTCAGGGCTTCATAAAATGCTGGCCCC 584  
  
RESULT 11  
BE205450  
LOCUS BE205450 583 bp mRNA linear EST 05-SEP-2000  
DEFINITION EST398126 KV0 Medicago truncatula cDNA clone pKV0-21N1, mRNA  
sequence.  
ACCESSION BE205450  
VERSION BE205450.1 GI:8748748  
KEYWORDS EST.  
SOURCE Medicago truncatula (barrel medic)  
ORGANISM Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
Medicago.  
REFERENCE 1 (bases 1 to 583)  
AUTHORS VandenBosch,K., Endre,G., Hur,J., Moore,J., Beremand,P., Ellis,L.,  
Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and  
Fraser,C.M.  
ESTs from uninoculated seedling roots of Medicago truncatula  
Unpublished (1999)  
CONTACT: VandenBosch K  
Department of Plant Biology  
University of Minnesota  
220 BioSci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA  
Tel: 612 624 2755  
Fax: 612 625 1738  
Email: kvanden@cbs.umn.edu  
Texas A&M University name:T265222e  
TIGR sequence name:MTGBM73TK  
More information is available at.  
http://chrysie.tamu.edu/medicago  
Seq primer: SKmod (CTA gAA CTA gtg gAT CC).  
  
FEATURES  
source  
1. .583  
/organism="Medicago truncatula"  
/mol\_type="mRNA"  
/cultivar="genotype A17"  
/db\_xref="taxon:3880"  
/clone="pKV0-21N1"  
/tissue\_type="Seedling roots"  
/dev\_stage="Immediately prior to inoculation with  
Sinorhizobium meliloti (0 hour)"  
/lab\_host="E.coli strain XL0LR"  
/clone\_lib="KV0"  
/note="Vector: pBluescript SK -; Site\_1: EcoRI; Site\_2:  
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA  
was directionally ligated into the Unizap XR vector from  
Stratagene and packaged using Gigapack III Gold packaging  
extracts. Plasmids containing cDNA inserts were excised  
from the recombinant lambda-zap phage using Ex-assist  
helper phage and propagated in XL0LR cells."

ORIGIN

Query Match 8.4%; Score 351.4; DB 10; Length 583;  
Best Local Similarity 77.4%; Pred. No. 1.6e-68;  
Matches 456; Conservative 0; Mismatches 121; Indels 12; Gaps 2;  
  
QY 914 CAATTCAAAGCATTCGCCCTTTTAGCTTTCACTTTTTGAGGAGAACCTC-----ATTCCGC 967  
Db 1 CTATACAAAAGTATTGCTTTCTTAGCATCCCTTTTGAAGAGGGTGTCTCTGCACATCCAC 60  
  
QY 968 ATTTGGTAATTGCTCCTCTATCGACTCTGCGTAACTGGGAGAGAGAGTTTGCCACATGGG 1027  
Db 61 ATCTGGTGGTTGCTCCACTTTCACACACTGCGAAACTGGGAACGTGAATTTGCAACATGGG 120  
  
QY 1028 CCCACAGATGAACGTGGTTATGTATTTTGGCACTGCGCAAGCTCGAGCAGTTATCAGAG 1087  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||



Db 121 CCCCTCAAAATGAATGTTATTATGTATGTTGGATCTGCCAAGCTCGTAGTGTATATCAGAG 180

QY 1088 AACATGAGTTTTTACTTATCGAAAGATCAAAAAAGATCAAGAAAAAGAAATCTGGACAAA 1147

Db 181 AATATGAATTTTACTTTCCCAAGAACTGAAGAAGAACAAAGAAAAAGAAATCT-----T 234

QY 1148 TAAGTAGCGAAAGCAAGCAAAAAAGAAATCAAGTTTGATGTCTCCTCACATCGTATGAGA 1207

Db 235 TAGTTAGTGAAGTAAGCATGACAGGATTAAGTTTGATGTCTCTTTTGACATCATATGAGA 294

QY 1208 TGATCAACCTAGATTTCAGCAGTTCTTAAACCAATTAAGTGGGAGTGCATGATTGTTGATG 1267

Db 295 TGATCAACCTTAGACACAAACATCATTTAAACCTATAAAATGGGAGTGCATGATAGTTGATG 354

QY 1268 AAGGTCATCGACTGAAAAATAAGGATTCAAGCTGTTCTCTTCAATTGACACAGTATTCAA 1327

Db 355 AAGGTCACCGCTCAAAAAATAAGGATTCAAAATATTATTTCTTCATTGAAGCAATATTCTA 414

QY 1328 GTAACCAACCGTATTCTTCTGACAGGAACACCACTTCAGAACAACTTGGATGAACCTTTTCA 1387

Db 415 CCAGACATCGTGTGCTCTTGACTGGAACCTCTCTTCAGAACAACTTGGATGAACCTTTTTA 474

QY 1388 TGCTCATGCAATTTCTTGATCGGGGAAGTTTGGAAGTTTGAGGAGTTCCAGGAGGAGT 1447

Db 475 TGCTTATGCAATTCCTTGATCGGGGAAGTTTGCAAGTTTAGAGGAATTCGAAGAAGAT 534

QY 1448 TCAAAGATATTAAATCAAGAGGAGCAGATCTCAAGGTTGCACAAAAATGTT 1496

Db 535 TCAAGGATATCAATCAAGAGGAACAGATTTCAAAGCTTCATAAAATGCT 583

RESULT 12

CA240279 778 bp mRNA linear EST 25-SEP-2003

LOCUS SCSBFL4066A03.g FL4 Saccharum officinarum cDNA clone SCSBFL4066A03

DEFINITION 5', mRNA sequence.

ACCESSION CA240279

VERSION CA240279.1 GI:35313723

KEYWORDS EST.

SOURCE Saccharum officinarum

ORGANISM

Saccharum officinarum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Saccharum.

REFERENCE

1 (bases 1 to 778)

Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.

The libraries that made SUCEST

Genet. Mol. Biol. 24 (1-4), 1-7 (2001)

COMMENT

Contact: Arruda P

Centro de Biologia Molecular e Engenharia Genetica

Universidade Estadual de Campinas

Caixa Postal 6010, 13083-970, Campinas SP, Brazil

Tel: 55 19 3788 1137

Fax: 55 19 3788 1089

Email: parruda@unicamp.br

Clone distribution: clone distribution information can be found through the Brazilian Clone Collection Center (BCCC) at <http://www.bcccenter.fcav.unesp.br>

Plate: 066 row: A column: 03

Seq primer: T7 Promoter Primer.

FEATURES

source

1. .778

/organism="Saccharum officinarum"

/mol\_type="mRNA"

/db\_xref="taxon:4547"

/clone="SCSBFL4066A03"

/lab\_host="DH10B"

/clone\_lib="FL4"

/note="Organ: Developed inflorescence and rachis (20cm-long); Vector: pSport1; Site\_1: SalI; Site\_2: NotI; An unidirectional cDNA library generated from [Developed inflorescence and rachis (20cm-long)]. cDNA was prepared from polyA+ mRNA using SuperScript Plasmid System Kit

ORIGIN

Query Match 8.0%; Score 336; DB 14; Length 778;

Best Local Similarity 68.9%; Pred. No. 5.2e-65;

Matches 507; Conservative 0; Mismatches 214; Indels 15; Gaps 3;

(Invitrogen). The double-strand cDNAs were fractionated

in a sepharose CL-2B 40cm-columns and fragments sizing

between 0.8 and 1.5 Kb were directionally cloned into the

vector. Details of each source of RNA and library

construction can be obtained at

<http://sucest.lad.ic.unicamp.br/public>"

QY 1911 GTACGAGCGAATTGATGGAAGGTTGGCGGAGCTGAGCGGCAATAACGCATAGATCGGTT 1970

Db 55 GTCCGAACGCATTGATGGCAAGATAAAGTGGTGTGAAAGGCAGATACGGATAGATCGCTT 114

QY 1971 CAATGCCAAAAAATCTTAACAAGTTTGTGTTTTTGTCTCTCCACAAGAGCTGGTGGCTTAGG 2030

Db 115 CAATGCTAAGAAATTCGACTAGGTTTGTCTTCTTCTTCTACAGAGCTGGTGGTCTGGG 174

QY 2031 AATAAATCTTGCAACGGCTGATACAGTAATCATTTATGACAGTGAATCCTCATGC 2090

Db 175 AATAAATTTGGCAACTGCAGATACTGTAATCATCTATGACAGTATTGGAACCCACATGC 234

QY 2091 TGATCTTCAAGCAATGGCTAGAGCTCATCGACTTGGCCAAACAAATAAGGTGATGATTTA 2150

Db 235 GGATTGCAAGCTATGGCAAGAGCTCATCGCTTAGGACAGACTAGCAAGGTGATGATATA 294

QY 2151 TAGGCTCATAAACCGAGGCACATTGAAGAAAGGATGATGCAATTGACTAAAAAGAAAAT 2210

Db 295 CCGGCTTGTTAGCCGAGGTACAATTGAGGAACGAATGATGCAAGCTTACAAAGAAAAAGAT 354

QY 2211 GGTTCTAGAGCATCTTGTGTTGGGAAAACCTC---AAAAACAAAAACATTAATCAGGAAGA 2267

Db 355 TTTATTGGAGCACTTAGTTGTTGGTTCGACTCACAAAAGCTAATAATGTCAATCAGGAGGA 414

QY 2268 GTTAGATGACATCATCAGGTATGGATCAAAGGAGCTTTTGTGCTAGTGAAGATGATGAAGC 2327

Db 415 GCTGGATGATATTATACGCTATGGATCAAAGGAGCTNTTTGTAGTACGAGAATGACGAA-- 472

QY 2328 AGGAAAGTCTGGAAAAATTCATTATGATGATGCGGCTATAGACAAATTGCTTGATCGTGA 2387

Db 473 -----TCTCGCCAAATTCATTACGACGAAGCTGCAATTGAGAGGTTGTTAGACCCGTGA 525

QY 2388 TCTCGTGGAGGCAGAGGAAGTCTCAGTGGATGATGAAGAGGAGAAATGGATTCTTAAAGGC 2447

Db 526 TCAAGTTGATGGTGATGAA---TCTGTGGAAGATGAAGAAAGATGAATTTCTTAAAGG 582

QY 2448 TTTCAGGTTGGCTAATTTTGAATATATAGATGAAAAATGAGGAGCAGCATTAGAGGCACA 2507

Db 583 ATTCAAGGTTGCAAACTTTGAATACATCGACGGGCANAGGCTCANGCAGATAGAGAGGA 642

QY 2508 GAGAGTCGCTGCTGAAAGCAAAATCTTCAGCAGGCAATTTCTGATAGAGCAAGTTATTGGGA 2567

Db 643 GGAGGCCCGCAGAAAGGCTGCAGCTGAGGCTGAAAACCTCTGAAAGATTAAACTATTGGGA 702

QY 2568 AGAGTTGTTAAAAGATATAATTTGAGCTGCACCAGGCTGAGGAGCTTAATGCTCTTGGAAA 2627

Db 703 TGAACATATNGAAGGATAGATACGATGTACAGAAAAGTTGAGGAACATACCTGCTATGGGAAA 762

QY 2628 AAGGAAGAGAAAGTCGC 2643

Db 763 AAGGAAAAGAACCCGC 778

RESULT 13

BQ834228

LOCUS

DEFINITION

BQ834228 Arabidopsis lyrata cDNA clone PlWB2-E11 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

BQ834228 348 bp mRNA linear EST 14-MAR-2003

AlEST0190 Arabidopsis lyrata Inflorescence pCMV-PCR Library

Arabidopsis lyrata cDNA clone PlWB2-E11 3', mRNA sequence.

BQ834228

BQ834228.1 GI:28951543

EST.

SOURCE Arabidopsis lyrata  
ORGANISM Arabidopsis lyrata  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
REFERENCE 1 (bases 1 to 348)  
AUTHORS Barrier,M., Bustamante,C.D., Yu,J. and Purugganan,M.D.  
TITLE Selection on rapidly evolving proteins in the Arabidopsis genome  
JOURNAL Genetics 163 (2), 723-733 (2003)  
MEDLINE 22505405  
PUBMED 12618409  
COMMENT Contact: Barrier M  
Department of Genetics  
North Carolina State University  
3513 Gardner Hall, Box 7614, Raleigh, NC 27695, USA  
Tel: 919 515 1761  
Fax: 919 515 1695  
Email: mbarrie@unity.ncsu.edu  
Plate: 1 row: J column: 22  
Seq primer: T3.  
FEATURES  
source  
1..348  
/organism="Arabidopsis lyrata"  
/mol\_type="mRNA"  
/cultivar="Karhumaki"  
/db\_xref="taxon:59689"  
/clone="PIWB2-E11"  
/tissue type="Inflorescence"  
/clone\_lib="Arabidopsis lyrata Inflorescence pCMV-PCR  
Library"  
/note="Vector: pCMV-PCR (Stratagene); Created using PCR  
Library Construction kit (Stratagene)"  
ORIGIN  
Query Match 7.4%; Score 308; DB 13; Length 348;  
Best Local Similarity 92.8%; Pred. No. 9.8e-59;  
Matches 323; Conservative 0; Mismatches 25; Indels 0; Gaps 0;  
QY 3657 GGGAGTGAGCTTTATTGAGGTTGATGATAAATGCTTGATGGACTTCCTAAGACTGATCC 3716  
Db 1 GGGAGTGAGCTTTATTGAAGTCGATGATAAATGCTTGATGGACTTCCTAAGACTGATCC 60  
QY 3717 TATCACTTCAGAAGAAATTATGGGGCTGCTGTTGACAAACCAAGCGCGGTCGAAAT 3776  
Db 61 CATCACTTCAGAAGAAATTATGGTGGCTGCTGTTGACAAACCAAGCGCGGTTGAAAT 120  
QY 3777 AGCTCAACATTATAACCAAGATGTGCAAACTTCCTTGATGAGAACCGCTCGGGAATCAGTCCA 3836  
Db 121 AGCTCAACATTATAACCAAGATGTGCAAAAGTTTTTAAATGAGAACCGCTCGTGATCACTCCA 180  
QY 3837 AGCATATGTAAACAACCAACCAACCGAGTACCAAGGTGAATGAGAGCTTCCGTGCACTCAA 3896  
Db 181 AGCATATGTAAACAACCAACCAACCAAGTACCAAGGTGAATGAGAGCTTCTGTGCACTCGA 240  
QY 3897 ATCTATCAATGGTAACATTAGAACGATCCTTTCGACTCCACTTGACCAATCAAAGTCACA 3956  
Db 241 ATCTATCAATGGTAACATTAGAACGATCCTTTCGACTCCACTTGACCAATCAAAGTCACA 300  
QY 3957 TGAAGACGACACCAAGCCAGACCTAAACAATGTTGAGATGAAGGACAC 4004  
Db 301 TGAAAACGACACCAAGCCAAACCTTAAACAATGTTGACATGAAGGACAC 348  
RESULT 14  
B29916/c  
LOCUS B29916 503 bp DNA linear GSS 13-OCT-1997  
DEFINITION T19D20TRB TAMU Arabidopsis thaliana genomic clone T19D20, genomic  
survey sequence.  
ACCESSION B29916  
VERSION B29916.1 GI:2515882  
KEYWORDS GSS.  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
REFERENCE 1 (bases 1 to 503)  
AUTHORS Rounsley,S.D., Kelley,J.M., Field,C.E., Craven,M.B., Adams,M.D. and  
Venter,J.C.  
TITLE Use of a BAC End Sequence Database To Identify Minimal Overlaps for  
Arabidopsis Genomic Sequencing  
JOURNAL Unpublished (1997)  
COMMENT Other\_GSSs: T19D20TFB  
Contact: Steve Rounsley  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: rounsley@tigr.org  
Seq primer: M13 Reverse  
Class: BAC ends  
High quality sequence stop: 503.  
FEATURES  
source  
1..503  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone="T19D20"  
/sex="hermaphrodite"  
/clone\_lib="TAMU"  
/note="Vector: BelOBACII; Site\_1: HindIII; Site\_2:  
HindIII; Produced by Rod Wing"  
ORIGIN  
Query Match 7.2%; Score 300; DB 28; Length 503;  
Best Local Similarity 100.0%; Pred. No. 6.7e-57;  
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 3878 AGAGCTTCCGTGCACCTCAAATCTATCAATGGTAAACATTAAACACATCCTTTCGATTACAT 3937  
Db 503 AGAGCTTCCGTGCACCTCAAATCTATCAATGGTAAACATTAAACACATCCTTTCGATTACAT 444  
QY 3938 CTGATCAATCCAAAGTCACATGAAGACGACACCAAGCCAGACCTTAAACAATGTTGAGATGA 3997  
Db 443 CTGATCAATCCAAAGTCACATGAAGACGACACCAAGCCAGACCTTAAACAATGTTGAGATGA 384  
QY 3998 AGGACACGGCCGAAGAAACAAACCCGTTAAGAGGTTGGCGTCGTCGATCTGAATGTGGTGG 4057  
Db 383 AGGACACGGCCGAAGAAACAAACCCGTTAAGAGGTTGGCGTCGTCGATCTGAATGTGGTGG 324  
QY 4058 AGGGAGAGGAGAACACATTGCTGAAGCTAGTGAAGTGTGGAAGTGTGATGTAAATAATGGAAGCCA 4117  
Db 323 AGGGAGAGGAGAACACATTGCTGAAGCTAGTGAAGTGTGGAAGTGTGATGTAAATAATGGAAGCCA 264  
QY 4118 AAGAAGAAGAGAGCCAAAGAACACATGGTCGTTGATTGACTCAACTGGTAAATCAAGATTC 4177  
Db 263 AAGAAGAAGAGAGCCAAAGAACACATGGTCGTTGATTGACTCAACTGGTAAATCAAGATTC 204  
RESULT 15  
CD812910  
LOCUS CD812910 673 bp mRNA linear EST 10-JUL-2003  
DEFINITION BN10.022P15F020211 BN10 Brassica napus cDNA clone BN10022P15, mRNA  
sequence.  
ACCESSION CD812910  
VERSION CD812910.1 GI:32494850  
KEYWORDS EST.  
SOURCE Brassica napus (rape)  
ORGANISM Brassica napus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.  
REFERENCE 1 (bases 1 to 673)  
AUTHORS Genoplante.

TITLE Genoplante, a major partnership french program in plant genomics  
JOURNAL Unpublished (2003)  
COMMENT Contact: Genoplante  
Genoplante  
93, rue Henri Rochefort 91025 EVRY CEDEX France  
Tel: 33 1 69 47 54 00  
Fax: 33 1 69 47 54 10

This sequence has been generated in the framework of the french  
plant genomics programme 'Genoplante' (<http://www.genoplante.com>)  
and <http://genoplante-info.infobiogen.fr>.

FEATURES  
source  
1..673  
/organism="Brassica napus"  
/mol\_type="mRNA"  
/cultivar="Jet neuf"  
/db\_xref="taxon:3708"  
/clone="BN10022P15"  
/tissue\_type="seed"  
/clone\_lib="BN10"

ORIGIN

Query Match	7.1%;	Score 298.2;	DB 14;	Length 673;
Best Local Similarity	81.5%;	Pred. No. 1.8e-56;		
Matches 400;	Conservative 0;	Mismatches 73;	Indels 18;	Gaps 4;
Qy 3642	GATCGCTGACACAGTGGGAGTGAGCTTTATTGAGGTTGATGATAAATGCTTGATGGACT	3701		
Db 1	GGTCACTGACACAGTGGGGCCAGCTTTATTGAGGTCGATGATAAATGCTTCAGGCACT	60		
Qy 3702	TCCTAGACTGATCCTATCACTTCAGAAATAATTATGGGGCTGCTTGACAAACCA	3761		
Db 61	TCCCAAGACTGAGCCCATCACTTCAGAAGAGATTATGGTTGCTGCTGACAAACCA	120		
Qy 3762	AGCGGGGTGAAATAGCTCAACATTATAACCAGATGTGCAAACTTCTTGATGAGAACGC	3821		
Db 121	AGAGCGGTTGAGATAGCCCCAACITTTATAACAAGATGTGCAAAATTTAGATGAGAACGC	180		
Qy 3822	TCGGGAATCAGTCCAAGCATATGTAAACAACCAACCAACCGATACCAAGGTGAATGAGAG	3881		
Db 181	TCGTGAATCTGTCCAAGCATATGTGAACAACCAACCGTCTAATGGCAAACTGGGCGGAGAG	240		
Qy 3882	CTTCCGTGCACTCAAATCTATCAATGGTAAACATTAAACAATCCTTTTCGATTACATCTGA	3941		
Db 241	CTTCCGTTCCATAGAATC-----TAACATTAGCAGAAATCCTTTCCGCTCCATCTGA	291		
Qy 3942	TCAATCCAAGTCACATGAAGACGACACCAAGCCAGACCTAAACAATGTTGAGATGAAGGA	4001		
Db 292	TCAACTCAAGTCTACTGAA--GACACCAAGCCGAACCTAAACAATGTTGAGAT---GGA	345		
Qy 4002	CACGGCCGAAGAAACAAACCCGTTAAGAGGTGGCGTCGTCGATCTGAATGTGGTGAGGG	4061		
Db 346	CTCGGTGAAGAAACGAAACCCCTTGAGAGCAAG---CGTCGATCTGAATGTGGTGAGGG	402		
Qy 4062	AGAGGAGAACATTGCTGAAGCTAGTGAAGTGTGATGTAAAAATGGAAGAAGCCAAAGA	4121		
Db 403	AGAGGAGAACATTGGAGAAGGTAGTGAAGTGTGATGTGAAGATGGAAGAAGCCAAATG	462		
Qy 4122	AGAAGAGAAGC	4132		
Db 463	AGAAGAGAAGC	473		



**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 14, 2004, 01:51:05 ; Search time 158 Seconds  
(without alignments)  
2474.973 Million cell updates/sec

Title: US-10-049-137-2  
Perfect score: 7187  
Sequence: 1 MSSSLVERLIRSDRKPVYNL.....VDVKMEEAKEEEKPKNMVVD 1384

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %		Length	DB ID	Description
		Match	Match			
1	7187	100.0	1384	4	AAY72755	Aay72755 Arabidops
2	7184	100.0	1384	4	AAY72756	Aay72756 Arabidops
3	7011.5	97.6	1357	4	AAY72787	Aay72787 Arabidops
4	2091	29.1	1912	6	AAE36110	Aae36110 Human chr
5	2091	29.1	1912	6	AAE36109	Aae36109 Human chr
6	2088	29.1	1911	2	AAR99534	Aar99534 Dermatomy
7	2083	29.0	1912	4	AAM39288	Aam39288 Human pol
8	2083	29.0	1944	6	AAE36107	Aae36107 Human chr
9	2067.5	28.8	1936	4	AAM41074	Aam41074 Human pol
10	2007	27.9	1470	5	ABP64883	Abp64883 Human pro
11	1939.5	27.0	1377	6	AAE36108	Aae36108 Human chr
12	1874.5	26.1	979	7	ADB64292	Adb64292 Human pro
13	1738.5	24.2	883	4	ABB71771	Abb71771 Drosophil
14	1621.5	22.6	1739	6	AAE36105	Aae36105 Human chr
15	1621.5	22.6	1739	6	AAE36106	Aae36106 Human chr
16	1565.5	21.8	1559	6	AAE36103	Aae36103 Human chr
17	1562.5	21.7	1709	6	AAE36104	Aae36104 Human chr
18	1562.5	21.7	1709	6	ABO07199	Abc07199 Human p53
19	1522.5	21.2	1883	4	ABB60038	Abb60038 Drosophil
20	1498	20.8	2703	4	ABB60074	Abb60074 Drosophil
21	1492.5	20.8	1967	6	AAE35342	Aae35342 Human col
22	1492.5	20.8	1967	6	AAE36114	Aae36114 Human chr
23	1461.5	20.3	1468	6	ABR52904	Abr52904 Protein s
24	1442.5	20.1	2897	6	ABR82557	Abr82557 Human chr
25	1422.5	19.8	2713	6	AAE36113	Aae36113 Human chr

26	1407.5	19.6	2131	6	AAE36112	Aae36112 Human chr
27	1374.5	19.1	1010	4	ABG09719	Abg09719 Novel hum
28	1260	17.5	2464	4	AAM78919	Aam78919 Human pro
29	1224.5	17.0	1057	5	AAU93152	Aau93152 Arabidops
30	1224.5	17.0	1057	7	ADD31011	Add31011 Plant yie
31	1157	16.1	730	4	AB94089	Ab94089 Human pro
32	1156	16.1	730	5	ABB85037	Abb85037 Pain regu
33	1146.5	16.0	655	4	ABG09856	Abg09856 Novel hum
34	1139	15.8	1120	6	ABR52907	Abr52907 Protein s
35	1135.5	15.8	1129	6	ABR52906	Abr52906 Protein s
36	1131	15.7	1027	4	ABB63954	Abb63954 Drosophil
37	1131	15.7	1027	7	ADB79883	Adb79883 Rat putat
38	1120	15.6	1359	6	ABR53192	Abr53192 Protein s
39	1097	15.3	1302	5	ABP73697	Abp73697 Candida a
40	1095.5	15.2	1703	6	ABR52985	Abr52985 Protein s
41	1086.5	15.1	1104	6	ABJ25604	Abj25604 Aspergill
42	1086.5	15.1	1353	6	ABJ26204	Abj26204 Aspergill
43	1040.5	14.5	1537	4	ABB66989	Abb66989 Drosophil
44	1040.5	14.5	1638	4	ABB61946	Abb61946 Drosophil
45	1039.5	14.5	551	5	ABB98113	Abb98113 CHD activ

ALIGNMENTS

RESULT 1	
AA72755	
ID	AA72755 standard; protein; 1384 AA.
XX	
AC	AA72755;
XX	
DT	31-MAY-2001 (first entry)
XX	
DE	Arabidopsis thaliana PKL protein.
XX	
KW	PKL; chromo domain; helicase domain; DNA binding domain; CHD3 protein;
KW	zinc finger domain; transgenic plant; chromatin remodelling factor; LEC1;
KW	pickle root; developmental identity; embryo development.
XX	
OS	Arabidopsis thaliana.
XX	
FH	Key Location/Qualifiers
FT	Domain 49..96
FT	/label= Zinc_finger_domain
FT	/note= "This_region is specifically referred in in claim
FT	8"
FT	Domain 115..151
FT	/label= Chromo_domain_I
FT	/note= "This_region is specifically referred in in claim
FT	7"
FT	Domain 191..227
FT	/label= Chromo_domain_II
FT	/note= "This_region is specifically claimed in in claim
FT	9"
FT	Domain 293..739
FT	/label= Helicase_domain
FT	/note= "This_region is specifically claimed in in claim
FT	7"
FT	Binding-site 1069..1095
FT	/label= DNA_binding_domain
FT	/note= "This_region is specifically claimed in in claim
FT	7"
XX	
PN	WO200114519-A2.
XX	
PD	01-MAR-2001.
XX	
PF	18-AUG-2000; 2000WO-US022725.
XX	
PR	20-AUG-1999; 99US-0149975P.
XX	
PA	(PURD ) PURDUE RES FOUND.
XX	

PI Ogas JP, Somerville CR;  
XX  
DR WPI; 2001-218433/22.  
DR N-PSDB; AAD02817.  
XX  
PT Regulating plant developmental identity comprises transforming a host  
PT cell with a nucleic acid encoding a PKL protein having at least one  
PT chromo domain, a helicase domain and a DNA binding domain.  
XX  
PS Claim 62; Page 67-76; 87pp; English.  
XX  
CC The invention relates to a method of regulating plant developmental  
CC identity which involves transforming a host cell with a nucleic acid  
CC encoding PKL (pickle) protein having at least one chromo domain, a  
CC helicase domain, a DNA binding domain and a zinc finger domain. This  
CC method is useful for transforming a host cell with PKL proteins for  
CC regulating developmental identity, such as for regulating the transition  
CC from embryonic to post-embryonic development. The PKL proteins act as  
CC chromatin remodelling factors to repress transcription of LEC1, a protein  
CC that plays a role in regulating embryo development. The method is also  
CC used for generating transgenic plants expressing PKL proteins. The  
CC present sequence is Arabidopsis thaliana PKL protein found in pickle  
CC root. PKL functions in regulating developmental identity in host cells.  
CC It is classified as CHD3 protein based on the presence of chromo domain,  
CC SNF2-related helicase/ATPase domain, DNA-binding domain and PHD zinc  
CC finger domain  
XX  
SQ Sequence 1384 AA;  
  
Query Match 100.0%; Score 7187; DB 4; Length 1384;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1384; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MSSLVERLIRSDRKPVYNLDDSDDDFVPKKDRTFEQVEAIVRTDAKENACQACGESTN 60  
Db 1 MSSLVERLIRSDRKPVYNLDDSDDDFVPKKDRTFEQVEAIVRTDAKENACQACGESTN 60  
  
QY 61 LVSCNTCTYAFHAKCLVPPLKDAVENWRCPECVSPPLNEIDKILDCEMRPTKSSEQSSD 120  
Db 61 LVSCNTCTYAFHAKCLVPPLKDAVENWRCPECVSPPLNEIDKILDCEMRPTKSSEQSSD 120  
  
QY 121 AEPKPIFVKQYLKWKGLSYLHCSWVPEKEFEQKAYKSNHRLKTRVNNFHRQMESFNNSD 180  
Db 121 AEPKPIFVKQYLKWKGLSYLHCSWVPEKEFEQKAYKSNHRLKTRVNNFHRQMESFNNSD 180  
  
QY 181 DFVAIRPEWTTVDRIILACREEDGELEYLVKYKELSYDECYWESESDISTFQNEIQRFKDV 240  
Db 181 DFVAIRPEWTTVDRIILACREEDGELEYLVKYKELSYDECYWESESDISTFQNEIQRFKDV 240  
  
QY 241 NSRTRRSKQVDHKNRPDRFQQFDHTPEFLKGLHPYQLEGLNFLRFSWSKQTHVILADEM 300  
Db 241 NSRTRRSKQVDHKNRPDRFQQFDHTPEFLKGLHPYQLEGLNFLRFSWSKQTHVILADEM 300  
  
QY 301 GLGKTIQSIALLASLFEENLI PHLVIAPLSTLRNWEREFATWAPQMNVMVYFGTAQARAV 360  
Db 301 GLGKTIQSIALLASLFEENLI PHLVIAPLSTLRNWEREFATWAPQMNVMVYFGTAQARAV 360  
  
QY 361 IREHEFYLSKDQKKIKKKSGQISSSESKQRIKFDVLLTSYEMINLDSAVLKPIKWECEMI 420  
Db 361 IREHEFYLSKDQKKIKKKSGQISSSESKQRIKFDVLLTSYEMINLDSAVLKPIKWECEMI 420  
  
QY 421 VDEGHRLKNKDSKLFSSLTQYSSNHRILLTGTPLONNLDELFLMHFLDAGKFGSLEEFQ 480  
Db 421 VDEGHRLKNKDSKLFSSLTQYSSNHRILLTGTPLONNLDELFLMHFLDAGKFGSLEEFQ 480  
  
QY 481 EEFKDIHQEEQISRLHKMLAPHLLRRVKQDVMDMPKPKELILRVDLSLQKEYYKAIFT 540  
Db 481 EEFKDIHQEEQISRLHKMLAPHLLRRVKQDVMDMPKPKELILRVDLSLQKEYYKAIFT 540  
  
QY 541 RNYQVLTKKGGAQISLNNIMMELRKVCCHPYMLEGVPVIHDANEAFKQLLESCGKLQLL 600  
Db 541 RNYQVLTKKGGAQISLNNIMMELRKVCCHPYMLEGVPVIHDANEAFKQLLESCGKLQLL 600

QY 601 DKMMVKLKEQGHRVLIYITQFQHMLDLLEDYCTHKKWQYERIDGKVGGAERQIRIDRFNAK 660  
Db 601 DKMMVKLKEQGHRVLIYITQFQHMLDLLEDYCTHKKWQYERIDGKVGGAERQIRIDRFNAK 660  
  
QY 661 NSNKFCFLSTRAGGLGINLATADTVIIYSDWNPHADLOAMARAHRLGQTNKVMYIRLI 720  
Db 661 NSNKFCFLSTRAGGLGINLATADTVIIYSDWNPHADLOAMARAHRLGQTNKVMYIRLI 720  
  
QY 721 NRGTEIERMMQLTKKKMVLEHLVVGKLTQNINQEEELDDIIIRYSGKELFASEDDEAGKSG 780  
Db 721 NRGTEIERMMQLTKKKMVLEHLVVGKLTQNINQEEELDDIIIRYSGKELFASEDDEAGKSG 780  
  
QY 781 KIHYYDDAIDKLLDRDLVEAEESVVDDEENGFLKAFKVANFEYIDENEAAALEAQRVAA 840  
Db 781 KIHYYDDAIDKLLDRDLVEAEESVVDDEENGFLKAFKVANFEYIDENEAAALEAQRVAA 840  
  
QY 841 ESKSSAGNSDRASYWEEELLKDKFELHQAABELNALGKRKRSRKQLVSIEDDDLAGEDEVSS 900  
Db 841 ESKSSAGNSDRASYWEEELLKDKFELHQAABELNALGKRKRSRKQLVSIEDDDLAGEDEVSS 900  
  
QY 901 DGDESYEAESTDGEAAGQGVQTRRYPYRRKGRDNLEPTPLMEGEGRSFVLGFNQSORAI 960  
Db 901 DGDESYEAESTDGEAAGQGVQTRRYPYRRKGRDNLEPTPLMEGEGRSFVLGFNQSORAI 960  
  
QY 961 FVQTLRMRYGAGNFDWKEFVPRLKQKTFEEINEYGILFLKHIAEEIDENSPTFSGVPKEG 1020  
Db 961 FVQTLRMRYGAGNFDWKEFVPRLKQKTFEEINEYGILFLKHIAEEIDENSPTFSGVPKEG 1020  
  
QY 1021 LRIEDVLVRIALLILVQEKVKFVEDHPGKPVFPSPRILERFPLGRSKIWKEEHDKIMIRA 1080  
Db 1021 LRIEDVLVRIALLILVQEKVKFVEDHPGKPVFPSPRILERFPLGRSKIWKEEHDKIMIRA 1080  
  
QY 1081 VLKHGYGRWQAI VDDKELGIQELICKELNFPPHISLSAAEQAGLQONGSGSNPGAQTNQ 1140  
Db 1081 VLKHGYGRWQAI VDDKELGIQELICKELNFPPHISLSAAEQAGLQONGSGSNPGAQTNQ 1140  
  
QY 1141 NPGSVITGNNNASADGAQVNSMFYIYRDMQRRLLVEFVKRVLILLEKAMNYEYAEYYGLGG 1200  
Db 1141 NPGSVITGNNNASADGAQVNSMFYIYRDMQRRLLVEFVKRVLILLEKAMNYEYAEYYGLGG 1200  
  
QY 1201 SSSIPTEEPEAEPKIADTVGVSVFIEVDDMDLGLPKTDPTITSEEINGAAVDNNQARVEIA 1260  
Db 1201 SSSIPTEEPEAEPKIADTVGVSVFIEVDDMDLGLPKTDPTITSEEINGAAVDNNQARVEIA 1260  
  
QY 1261 QHYNQMCKLLDENARESVOAYVNNQPPSTKVNESFRALKSINGNINTILSITSDQSKSHE 1320  
Db 1261 QHYNQMCKLLDENARESVOAYVNNQPPSTKVNESFRALKSINGNINTILSITSDQSKSHE 1320  
  
QY 1321 DDTKPDLLNNVEMKDTAEETKPLRGVVDLNNVVEGEENIAEASGSVDVKMEEAKEEEKPKN 1380  
Db 1321 DDTKPDLLNNVEMKDTAEETKPLRGVVDLNNVVEGEENIAEASGSVDVKMEEAKEEEKPKN 1380  
  
QY 1381 MVVD 1384  
Db 1381 MVVD 1384  
  
RESULT 2  
AA72756  
ID AA72756 standard; protein; 1384 AA.  
XX  
AC AA72756;  
XX  
DT 31-MAY-2001 (first entry)  
XX  
DE Arabidopsis thaliana mutant PKL protein (K304R).  
XX  
KW PKL; chromo domain; helicase domain; DNA binding domain;  
KW zinc finger domain; transgenic plant; chromatin remodelling factor; LEC1;  
KW mutant; mutein; pickle; developmental identity; embryo development.  
XX  
OS Arabidopsis thaliana.  
OS Synthetic.







Db 1174 SSSIPTPEAEPIADTVGVSFIEVDDEMLDGLPKTDPITSEEIMGAAVDNNQARVEIA 1233

Qy 1261 QHYNQMCKLLDENARESVOAYVNNQPPSTKVNESFRALKSINGNINTILSITSDQSKSHE 1320

Db 1234 QHYNQMCKLLDENARESVOAYVNNQPPSTKVNESFRALKSINGNINTILSITSDQSKSHE 1293

Qy 1321 DDTKPDNNVEMKDTAEETKPLRGVVDLNVVEGEENIAEASGSDVKMEEAKEEEKPKN 1380

Db 1294 DDTKPDNNVEMKDTAEETKPLRGVVDLNVVEGEENIAEASGSDVKMEEAKEEEKPKN 1353

Qy 1381 MVVD 1384

Db 1354 MVVD 1357

RESULT 4

AAE36110

ID AAE36110 standard; protein; 1912 AA.

XX AC AAE36110;

DT 26-JUN-2003 (first entry)

DE Human chromodomain helicase DNA binding protein (CHD) #8.

KW Human; p53 pathway; therapeutic; angiogenic disorder; apoptotic disorder; chromodomain helicase DNA binding protein; CHD; cancer; gene therapy; cell proliferative disorder; chromatin organisation modifier domain; cytostatic.

OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Domain 619..661

FT /note= "Chromodomain"

FT Domain 729..1025

FT /note= "SNF2N domain"

FT Domain 1090..1164

FT /note= "Conserved C-terminal domain"

XX WO200298899-A2.

PD 12-DEC-2002.

PF 03-JUN-2002; 2002WO-US017466.

XX 05-JUN-2001; 2001US-0296076P.

PR 10-OCT-2001; 2001US-0328605P.

PR 22-OCT-2001; 2001US-0338733P.

PR 15-FEB-2002; 2002US-0357253P.

PR 15-FEB-2002; 2002US-0357600P.

XX (EXEL-) EXELIXIS INC.

XX FI Friedmann L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP; Lioubin MN;

XX WPI; 2003-156840/15.

PT Identifying a candidate p53 pathway-modulating agent as therapeutic targets for disorders related to defective p53 function e.g. cancer by contacting an assay system having purified CHD polypeptide or nucleic acid, with a test agent.

PS Claim 13; Page 229-237; 278pp; English.

CC The present invention relates to a method for identifying candidate p53 pathway modulating agents. The method involves contacting an assay system comprising purified chromatin organisation modifier (chromo) domain helicase DNA binding proteins (CHD), nucleic acids, their functionally active fragments or derivatives, with a test agent under conditions where, but for the presence of the test agent, the system provides a reference activity. The methods are useful for identifying modulators of

CC the p53 pathway as therapeutic targets for disorders associated with defective p53 function, such as angiogenic disorders, apoptotic disorders or cell proliferative disorders, e.g. cancer. The modulators are useful as research reagents, diagnostics and therapeutics. The invention is also useful in gene therapy. The present sequence is human chromodomain helicase DNA binding protein

XX SQ Sequence 1912 AA;

Query Match 29.1%; Score 2091; DB 6; Length 1912;

Best Local Similarity 35.5%; Pred. No. 4.3e-149;

Matches 520; Conservative 205; Mismatches 424; Indels 314; Gaps 44;

Qy 21 DSDDDDFVPKKDRTFEQVEAIVRTDAKENACQAGESTNLVSCNTCTYAFHAKCLVPPL 80

Db 426 DNSEGEILEEVGDLEED-----DHHMEFCRVCKDGGELCCDTCPSSYHIHCLNPPL 480

Qy 81 KDASVENWRCPECVSPL--NEIDKILDCE--MRPTKSSEQSSDAB-----PKPIF--V 128

Db 481 PEIPNGEWLCPRCTCPALKGVQKILIKWKGPPPTVPVPRPDADPTSPKPLEGRPE 540

Qy 129 KQYLVKWKGSLYHCSWVPEK-----FQKAYKSNHRLKTRVNNFHRQME----- 173

Db 541 RQFFVKQGMYSYWHCSWVSELQLELHCQVMFRNYQRKNDMDPEPSGDFGGDEKSRKRN 600

Qy 174 ---SPNSEDDF--VAIRPEWTTVDRIL-ACREEDGELEYLVKYKELSYDECYWESE--- 224

Db 601 KDPKFAEMEERFYRIGIKPEWMMIHRILNHSVDKKGHVHYLIKWRDLPYDQASWESEDVE 660

Qy 225 -SDISTFQNEIQRFKDV-----NSRTRRSKVDHKNR-----DFQQFDHTPEFLK 270

Db 661 IQDYDLFKQSYNNHRELMRGEEGPGKKLKKVLRKLERPPETPTVPTVKYERQPEYLD 720

Qy 271 ---GLLHPYQLEGLNFLRFSKQTHVILADEMGLKTIQSIALLASLFEE--NLIPHVLV 325

Db 721 ATGGTLHPYQMEGLNWLRFSAQGTDTILADEMGLKTVQTAFLYSLYKEGHSKGPFVLV 780

Qy 326 IAPLSTLRNWEREFATWAPQMVVMYFGTAQARAVIREHEFVLSKDQKIKKKSGQISS 385

Db 781 SAPLSTIINWEREFEMWAPDMYVTVYVGDKDSRAIIRENEFSFEDNAIRGGKASRM--- 837

Qy 386 ESKQKRIKFDVLLTSYEMINLDSAVLKPIKWECEMIVDEGHRKKNKDSKLFSSLTQYSSNH 445

Db 838 -KKEASVKFHVLLTSYELITIDMAILGSIDWACLIVDEAHLKNNQSKFFRVNLNGYSLOH 896

Qy 446 RIILGTGPLQNNLDELFMHFLDAGKFGSLEEFQEEFKDINQEEQISRLHKMLAPHLLR 505

Db 897 KLLLTGTPLQNNLEELFHLNFLTPTPERFHNLEGFLEEFADIAKEDIKKLHMLGPHMLR 956

Qy 506 RVKQDVWKDMPKPKELILRVLDLSSLOKEYYKAIFTRNYQVLTKKGGA-QISLNNIMMELR 564

Db 957 RLKADVFKNMPSKTELIVRVELSPMQKKYKYILTRNFEALNARGGNGVSLNVVMDLK 1016

Qy 565 KVCCHPYMLEGVEPVIHDANEAFKQ-----LLESCGKLQLLDKMMVKLKEQGHVRL 615

Db 1017 KCCNHPYLF----PVA--AMEAPKMPNGMYDGSALIRASGKULLLLQKMLKNLKEGGHRLV 1070

Qy 616 IYTFQHMULDLEDYCTHKWQYERIDGKVGGAEQRIIDRFNAKNSNKFCLLLSTRAGG 675

Db 1071 IFSQMTKMLDLEDLEFHEGKYVERIDGGITGNMRQEAIDRFNAPGAQFCFLLLSTRAGG 1130

Qy 676 LGINLATADTVIIYDSDNPHADLQAMARAHRLGQTNKVMYRLINRGTEIERMMQLTKK 735

Db 1131 LGINLATADTVIIYDSDNPHNDIQAFSRAHRIQNKVMYRFBVTRASVEERITQVAKK 1190

Qy 736 KMVLEHLVVGK--LKTQINQEEELDDIIRYGSKELFASEDEAGKSGK-----IHYD 785

Db 1191 KMMLTHLVVRPGLSGKSGTSGMSKQELDDILKFGTEELFKDEATDGGGDNKEGEDSSVIHYD 1250

Qy 786 DAAIDKLLDRDLVEABEVSVDDEENGFLKAFKVANFEYIDENEAAALEAQRVAESKSS 845

Db 1251 DKAIERLLDRNQDETEDEL--QGMNEYLSSFKAQYVVREEMGEEMEEVEREIIKQEES 1308





Db	1017	KCCNHPYLF-----PVA--AMEAPKMPNGMYDGSALIRASGKLLLLQKMLKNLKEGHRVL	1070
Qy	616	IYTOFQHMLDLLEDYCTHKWQYERIDGKVGGAERQIRIDRFNAKNSNKFCFLSTRAGG	675
Db	1071	IFSQMTKMLDLLEDYCTHKWQYERIDGKVGGAERQIRIDRFNAKNSNKFCFLSTRAGG	1130
Qy	676	LGINLATADTVIYDSDNWPHADLQAMARAHRLGQTNKVMYRLNRGTIERMMQLTKK	735
Db	1131	LGINLATADTVIYDSDNWPHNDIQAFSRAHRIQGNKKVMYRFTTRASVEERITQVAKK	1190
Qy	736	KWVLEHLVVGK---LKTQINQOEELDDIIRYGSKELFASDEDEAGKSGK-----IHVD	785
Db	1191	KWVLEHLVVGK---LKTQINQOEELDDIIRYGSKELFASDEDEAGKSGK-----IHVD	1250
Qy	786	DAIDKLLDRDLVEABEVSVDDEENGELKAPKVANFEYIDENEAALAEQRAVAESKSS	845
Db	1251	DAIERLLDRNQDETEDEL--QGMNEYLSFFKVAQYVVREEMGEEMEEVEREIIKQEES	1308
Qy	846	AGNSDRASYWEELLKDKFELHQAEEELNALGKRKRKQL----VSIEEDDLAGLEDVSSD	901
Db	1309	VD----PDYWEKLLRHHYEQQEDLARNLGKGRIRKQVNYNDGQEDRD---WQDDQSD	1361
Qy	902	GDESYEASTDG--EAAGQGVQTGRRPYR--RKGRDNLEPTPLMEGERSFRVLGFNQSQ	957
Db	1362	NQSDYSVASEGDEDFDERSEAPRRPSRKLNRDKKPLP-PLLARVGNIEVLGFNARQ	1420
Qy	958	RAIFVQTLNRYG-----AGNFDWKEFVPRLLKQKTFEEINEYIGILFLKHIAEIDENSPTE	1012
Db	1421	RKAFLNAINRYGMPPODAFTTQW--LVRDLRGKSEKEFKAYVSLFMRHLCEPGADGAET	1478
Qy	1013	SDGVPKEGLRIEDVLRIALLILVQEKVFEVDHPGKPVFPSPRILERFPGLRSKIKWEE	1072
Db	1479	ADGVPREGLSRQHVLTRIGVMSLRIRKVKQEF-----	1510
Qy	1073	HDKIMIRAVLKHGYGRWQAIVDVKELGELQELICKELNFPHISLSAAEAGLQONGSGGS	1132
Db	1511	-----HVNGRW-----SMP--ELAEVEENKMSQPGS--P	1536
Qy	1133	NPQAQTNQNPQSVITGNNNASADGAQVNSMFYRDMQRRLLVEFKRVLLLEKAMNVEYA	1192
Db	1537	SPKTPTPTPTPD-----TQPTI-----	1553
Qy	1193	EYYGLGGSSSIPTEEPEAEPKIADTVGVSVFIEVDDEMLDGLPKTDPITSEEIMGAADV	1252
Db	1554	-----PAPVPPAE-----DGIK-----	1565
Qy	1253	NQARVEIAQHYNQMKLLDENARESVOAYVNNQPPSTKVNESFRALKSINGNI-NTILSI	1311
Db	1566	-----IEEN-----SLKEEESIEGEKEVKSTAPETAIEC	1594
Qy	1312	TSDQSKSHEDT-----KPDLLNNVEMKDTAE---ETKPLRGVVDLNVVEGEENIAE	1360
Db	1595	TQAPAPASEDEKVVVEPPEGEKEKVEKAEVKERTPEETEP--KGAADVEKVE-EKSAID	1651
Qy	1361	ASGSVDVKMEAEKEEKPKNMVV	1383
Db	1652	LTPIVVEDKEKEKEEKEVML	1674
RESULT 6			
AAR99534			
ID	AAR99534	standard; protein; 1911 AA.	
XX	AC	AAR99534;	
XX	DT	30-OCT-1996 (first entry)	
XX	DE	Dermatomyositis specific autoantigen, Mi-2.	
XX	KW	Mi-2; autoantigen; collagen disease; chromosome 12; 12p13; helicase;	
XX	KW	dermatomyositis; diagnosis.	
OS	OS	Homo sapiens.	

XX	Key	Location/Qualifiers
FT	Region	53. .73
FT		/label= Region_a
FT		/note= "contains 3 potential core target motifs"
FT	Region	113. .133
FT		/label= Region_b
FT		/note= "contains 4 potential core target motifs"
FT	Region	133. .143
FT		/note= "possible electrostatic interaction with chromatin of histones"
FT	Region	257. .287
FT		/label= Region_c
FT		/note= "contains 3 potential core target motifs"
FT	Region	747. .758
FT		/note= "Helicase-specific motif I"
FT	Region	782. .793
FT		/note= "Helicase-specific motif IA"
FT	Region	869. .877
FT		/note= "Helicase-specific motif II"
FT	Region	897. .911
FT		/note= "Helicase-specific motif III"
FT	Region	943. .959
FT		/label= Region_d
FT		/note= "contains 1 potential core target motif"
FT	Region	949. .960
FT		/note= "Helicase-specific motif IV"
FT	Region	1121. .1144
FT		/note= "Helicase-specific motif V"
FT	Region	1149. .1183
FT		/note= "Helicase-specific motif VI"
XX	DE19509279-C1.	
PN		
XX	15-MAY-1996.	
PD		
XX		
XX	15-MAR-1995; 95DE-01009279.	
PF		
XX	15-MAR-1995; 95DE-01009279.	
PR		
XX	(PRIV-) PRIVATES INST IMMUNOLOGIE & MOLEKULARGEN.	
PA	Seelig HP, Renz M;	
XX		
PI		
XX		
DR	WPI; 1996-240280/25.	
DR	N-PSDB; AAT32301.	
XX	DNA encoding dermatomyositis specific auto:antigen - useful for differential diagnosis and treatment of dermatomyositis.	
PT	Claim 1; Fig 2; 20pp; German.	
XX		
PS	The present sequence is that of a 218 kD dermatomyositis specific auto-antigen, designated Mi-2. The sequence numbering given in the specification starts at amino acid 2, i.e. the first Met residue is omitted. The protein is hydrophilic, acidic and protruding regions of the protein are characteristic of helicases. The gene corresponding to the cDNA (AAT32301) encoding Mi-2, was localised to chromosome 12 (12p13).	
CC	The DNA can be used for the recombinant production of Mi-2 which is used for, e.g. the differential diagnosis of collagen diseases, esp. dermatomyositis, e.g. by immunoassay or Western blotting	
XX		
SQ	Sequence 1911 AA;	
Query Match 29.1%; Score 2088; DB 2; Length 1911;		
Best Local Similarity 35.5%; Pred. No. 7.3e-149;		
Matches 519; Conservative 206; Mismatches 424; Indels 314; Gaps 44;		
Qy	21	DSDDDDDFVPKKDRTFEQVBAIVRTDAKENACQACGESNLVSCNTCTYAFHAKCLVPPL 80
Db	425	DNSEGEIILEVGGDLEED----DHHMEFCRVCKDGGELLCCDTCPSSYHIHCLNPPL 479
Qy	81	KDASVENWRCPECVSPL--NEIDKILDCE--MRPTKSSEQSSDAE-----PKPIF---V 128

Db 480 PEIPNGEWLCPRCTCPALKGVQKILIKWGWQPPSPPTVPVRPPDADNTSPSPKPLEGRPE 539  
QY 129 KOYLKWKGLSYLHCSWVPEKE-----FQKAYKSNHRLKTRVNNFHRQME----- 173  
Db 540 RQFFVKWQGSYWHCSWVSELQLELHCQVMFERNYQRKNDMDPEPPSGDFGDEEKSRRKN 599  
QY 174 ---SFNSENDDF--VAIRPEWTTVDRI--ACREEDGELEYLVKYKELSYDECYWESE--- 224  
Db 600 KDPKFAEMEERFYRYGIKPEWMMIHRILNHSVDKKGHVHYLIKWRDLPYDAQSWESESDVE 659  
QY 225 -SDISTFQNEIQRFKDV-----NSRTRRSKDVVDHQRNPR-----DFQQFDHTPEFLK 270  
Db 660 IQDYDLFKQSYWNHRELMRGEGRPGKKLKKVKLRKLERPPETPTVDPTVKYERQPEYLD 719  
QY 271 ---GLLHPYQLBGLNFLRFSWSKQTHVILADEMGLGKTIQSIALLASLFEE--NLIPHVL 325  
Db 720 ATGGTLHPYQMEGLNWLRFWSAQGTDTILADEMGLGKTQTAVFLYSLYKEGHSKGPFVL 779  
QY 326 IAPLSTLRNWEREFATWAPQMNVMYFGTAQARAVIREHEFYLSKDQKIKKKSGQISS 385  
Db 780 SAPLSTIINWEREFEMWAPDMYVVTVYVGDKSRAIIRENEFSFEDNAIRGGKASRM--- 836  
QY 386 ESKQKRIKFDVLLTSYEMINLDSAVLKPWKECMIVDEGHRLLKNKDSKLFSSLTQYSSNH 445  
Db 837 -KKEASVKFHVLLTSYELITIDMAILGSDWACLIVDEAHRLLKNQSKFFRVLNGYSLOH 895  
QY 446 RILLTGTPLONNDELFLMLHFLDAGKFSLEEFQEEFKDINQEEQISRLHKLAPHLR 505  
Db 896 KLLLTGTPLQNNLEELFHLNFLTPTPERFHNLEGFLEEFADIAKEDQIKKLHMLGPHMLR 955  
QY 506 RVKQDVMDMPKPKELILRVDLSSLOKEYYKAIFTRNYQVLTKKGA-QISLNNIMMELR 564  
Db 956 RLKADVFKMNPSTELIVRVELSPMQKYYKYVILTRNFEALNARGGQVSLNVMVDLK 1015  
QY 565 KVCCHPYMLEGVEPVIHDANAEAFKQ-----LLESCGKLQLLDKMMVKLKEQGHRLV 615  
Db 1016 KCCNHPYLF----PVA--AMEAPKMPNGMYDGSALIRASGKLLLLQKMLKNLKEGGHRLV 1069  
QY 616 IYTFQHMULDLEDYCTHKKWQYERIDGKVGAERQIRIDRFNAKNSNKFCFLSTRAGG 675  
Db 1070 IFSQMTKMLDLEDLEHEGYKYERIDGGITGNMRQEAIDRFNAPGQOQFCFLSTRAGG 1129  
QY 676 LGINLATADTVIYSDWNPHADLOAMARAHRLGQTNKVMYRLNRTGIEERMQLTKK 735  
Db 1130 LGINLATADTVIYSDWNPHNDIQAFSRAHRIQONKKVMYRFVTRASVEERITQVAKK 1189  
QY 736 KMVLEHLVVGK---LKTQINQEEELDDIIRYGSKELFASEDDEAGKSGK-----IHYD 785  
Db 1190 KMMLTHLVVRPGLGSKTGSMSKQELDDILKPGTEELFKDEATDGGDNKEGEDSSVIHYD 1249  
QY 786 DAAIDKLLDRDLVEABEVSVDDEEENGFLKAFKVANFEYIDENEAALAEQAVAAESKSS 845  
Db 1250 DKAIERLLDRNQDETDTEL--QGMNEYLSSPFKVAQYVVREEMGEEDVEREIKQEES 1307  
QY 846 AGNSDRASYWEELLKDKFELHQAEEELNALGKRKRSRKQL---VSIERDDLAGLEDVSSD 901  
Db 1308 VD----PDYWEKLLRHYYEQQEDLARNLKGKRIKQVNYNDGSEQEPRD---WQDDQSD 1360  
QY 902 GDESIEAESTDG--BAAGQGVQGTGRRPYR---RKGRDNLEPTPLMEGEGRSFRVLGFNQSQ 957  
Db 1361 NQSDYSVASEEGDEDFDERSEAPRRPSRKGLRNDKDKPLP-PLLARVCGNIEVLGFNARQ 1419  
QY 958 RAIFVQTLMRYG-----AGNFDWKEFVPRLKQKTFFEEINEYIGILFKHIAEIEDNSPTF 1012  
Db 1420 RKAFLNAIMRYGMPFPQDAFTTQW--LVRLDRGKSEKEFKAYVSLFMRHLCPEPGADGETF 1477  
QY 1013 SDGVPKEGLRIEDVLVRIALLILVQEKVKFVEDHPGKPVFPSPRILERPFLRSWKIWKEE 1072  
Db 1478 ADGVPREGLSRQHVLTRIGVMSLIRKKVQEFE-----1509  
QY 1073 HDKIMIRAVLKHGGRWQAIVDDKELGIQELICKELNFPFHISLSAAEQAGLQONGSGGS 1132

Db 1510 -----HVNGRW-----SMP--ELAEEVENKMSQPGS--P 1535  
QY 1133 NPGAQTNQNPQSVITGNNNASADGAQVNSMFYFRDMQRRLVEFVKRVLLEKAMNYEYA 1192  
Db 1536 SPKTPPTSTPGD-----TQPNT----- 1552  
QY 1193 EEYIGLGGSSSIPTEEPEAEPKIADTVGSFIEVDDEMLDGLPKTDPTITSEEIMGAAVDN 1252  
Db 1553 -----PAPVPPAE-----DGIK----- 1564  
QY 1253 NQARVEIAQHYNQCKLLDENARESQQAYVNNQPPSTKVNESFRALKSINGNI-NTILSI 1311  
Db 1565 -----IBEN-----SLKEEESIEGEKEVKSTAPETAIEC 1593  
QY 1312 TSDQSKSHEDDT-----KPDINNVMKDTAE---ETKPLRGGVVDLNVVEGEENIAE 1360  
Db 1594 TQAPAPASEDEKVVVEPPEGEKEKAEVKERTEEPMETEP--KGAADVEKVE-EKSAID 1650  
QY 1361 ASGSVDVKMEEAKEEEKPKNMVV 1383  
Db 1651 LTPIVVEDKBEKKEEEKKEVML 1673  
RESULT 7  
AAM39288  
ID AAM39288 standard; protein; 1912 AA.  
XX  
AC AAM39288;  
XX  
DT 22-OCT-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 2433.  
XX  
KW Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia.  
XX  
OS Homo sapiens.  
XX  
PN WO200153312-A1.  
XX  
PD 26-JUL-2001.  
XX  
PF 26-DEC-2000; 2000WO-US034263.  
XX  
PR 23-DEC-1999; 99US-00471275.  
PR 21-JAN-2000; 2000US-00488725.  
PR 25-APR-2000; 2000US-00552317.  
PR 20-JUN-2000; 2000US-00598042.  
PR 19-JUL-2000; 2000US-00620312.  
PR 03-AUG-2000; 2000US-00653450.  
PR 14-SEP-2000; 2000US-00662191.  
PR 19-OCT-2000; 2000US-00693036.  
PR 29-NOV-2000; 2000US-00727344.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
PI Zhou P, Goodrich R, Drmanac RT;  
XX  
DR WPI; 2001-442253/47.  
DR N-PSDB; AAI58444.  
XX  
PT Novel nucleic acids and polypeptides, useful for treating disorders such  
PT as central nervous system injuries.  
XX  
PS Example 4; SEQ ID NO 2433; 10078pp; English.  
XX  
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the



CC encoded polypeptides (AAM38642-AAM42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders. Note: The sequence data for this patent did not form  
CC part of the printed specification  
XX  
SQ Sequence 1912 AA;

Query Match 29.0%; Score 2083; DB 4; Length 1912;  
Best Local Similarity 35.7%; Pred. No. 1.8e-148;  
Matches 522; Conservative 201; Mismatches 425; Indels 316; Gaps 45;

QY	21	DDSDDDDFVPKDKORTFEQVEAIVRTDAKENACQACGESTNLVSCNTCTYAFHAKCLVPPL	80
Db	426	DNSEGEIIEVGGDLEED-----DHMEFCRVCKOGGELCCDTCSPSYHICLNPP	480
QY	81	KDASVENWRCECVSPL--NEIDKILDCB--MRPTKSSEQSSDAE-----PKPIF--V	128
Db	481	PEIPNGEWLPCRCTCPALKGKVQKILIKWQGPSPPTVPRPPDADNTSPKPLEGRPE	540
QY	129	KOYLKWKGLSYLHCSWVPEKE-----FQKAYKSNHRLKTRVNNFHRQME-----	173
Db	541	RQFFVKWQMSYWHCSWSVSELQLELHCQVMFERNYQKNDMDPEPSPGDFGDEEKSRRKN	600
QY	174	---SFNNSDDF--VAIRPEWTTVDRIIL--ACREEDGELEYLVKVKELSYDECYWESE---	224
Db	601	KDPKFAEMEERFYRYGIKPEWMMIHRILNHSVDKKGHVHYLIKWRDLPYDQASWESE	660
QY	225	-SDISTFQNEIQRFKDV--NSRTRRSKDVHDKRNPDRFQQFDHTP-----EFL	269
Db	661	IQDYDLFKQSYWNHRELMRGEEGRPGKKL--KKVKLRKLERPPETPTVDPTVKYERHAEYL	719
QY	270	K---GLLHPYQLEGLNFLRFSWSKQTHVILADEMGLKTIQSIALLASLFEF--NLIPHL	324
Db	720	DATGGTLHPYQMEGLNWLRFSAQGTDTILADEMGLKTVQTAFLYSLYKEGHSKGPF	779
QY	325	VIAPLSTLRNWEREFATWAPQNMVVMYFCTAQAARAVIREHEFYLSKDQKIKKKSGQIS	384
Db	780	VSAPLSTIINWEREFEMWAPDMYVVTVYVGDKDSRAIIRENEFSFEDNAIRGGKASRM--	837
QY	385	SESKQKRIKFDVLLTSYEMINLDSAVLKPIKWECEMIVDEGHRLLKNKDSKLFSSLTQYSSN	444
Db	838	--KKEASVKFHVLLTSYELITIDMAILGSDWACLIVDEAHLKKNQSKFFRVNLGYSIQ	895
QY	445	HRILLTGTPLONNLDELFLMLMHFLDAGKFGSLEEFQEEFKDINQEEQISRLHKMLAPHLL	504
Db	896	HKLLLTGTPLQNNLEELFHLNFLTPTPERFHNLEGFLEEFADIAKEDQIKKJHMLGPHML	955
QY	505	RRVKQVMKDMPPKKELILRVDSLSSQKEYYKAIFTRNYQVLTKKGA-QISLNNIMMEL	563
Db	956	RLKADVFNKMPSKTELIVRVELSPMQKYYKYILTRNFEALNARGGQVSLNVMVMDL	1015
QY	564	RVCCCHPYMLEGVEPVIHDANAEAFKQ-----LLESCGKLQLLDKMMVKLKEQGHV	614
Db	1016	KKCCNHPYLF---PVA--AMEAPKMPNGMYDGSALIRASGKLLLLQKMLKNLKEGGHRV	1069
QY	615	LIYTFQFQHMLDLLEDYCTHKKQYERIDGKVGGAERQIRIDRFNAKNSNKFCLLSTRAG	674
Db	1070	LIFSQMTKMLDLEDFLEHEGYKYERIDGGITGNMRQEAIDRFNAPGAQQAQFCFLSTRAG	1129
QY	675	GLGINLATADTVIYDSWNPHADLQAMARAHLRGQTNKVMYRLINRGITIEERMQLTK	734
Db	1130	GLGINLATADTVIYDSWNPNDIQAFSRAHRIGQNKKVMYRFFVTRASVEERITQVAK	1189

QY	735	KQWLEHLVVGK---LKTONINQOEELDDIIRYSGSKELFASDEAGSKGK-----IHY	784
Db	1190	KQWMLTHLVVRPGLSGKTSMSKQELDDILKFGTEELFKDEATDGGDNKEGEDSSVIHY	1249
QY	785	DDAIDKLLDRDLVEAEVSVDDDEENGFLKAFKVANFEYIDENEAAALEAQVAAESKS	844
Db	1250	DDKAIERLLDRNQDETEDEL--QGMNEYLSFFKVAQYVVREEMGEVEEVEIIKQEE	1307
QY	845	SAGNSDRASYWEELLKDKFELHQAELNALGKRKRSRKQL-----VSIEDDDL	900
Db	1308	SVD-----PDYWEKLLRRHHYEQQEDLARNLGKGRIRKQVNVNDGSQEDRD---WQDDQS	1360
QY	901	DGDESYEAESTDG-EAAGQGVQGTGRRPYR---RKGRDNLEPTPLMEGEGRSFRVLGFNQ	956
Db	1361	DNQSDYSVASEEGDEDFDERSEAPRRPSRKGLRNDKDKPLP-PLLARVGGNIEVLGFNAR	1419
QY	957	QRAIFVQTLTRYG-----AGNFDWKEFVPRLLKQKTFFEEINEYGIILFKHIAE	1011
Db	1420	QRKAFNLNAIMRYGMPDQDAFTTQW--LVRDLRGKSEKEFKAYVSLFMRHLCEPGADGAET	1477
QY	1012	FSDGVPKEGLRIEDVLVRIALLILVQEKVKFVEDHPGKVPFPRILERFPGLRSGKIWKE	1071
Db	1478	FADGVPREGLSRQHVLTIRIGVMSLIRKKVQEFE-----	1510
QY	1072	EHDKIMIRAVLKHGYGRWQAVDDKELGIELICKELNFPFHSLSAAEQAGLQONGSGG	1131
Db	1511	-----HVNGRW-----SMP--ELAEVEENKQMSQPGS--	1535
QY	1132	SNPGAQTNQNPQSVITGNNNASADGAQVNSMFYRDMQORLVEFVKRVLLEKAMNYEY	1191
Db	1536	PSPKTPTPSTPGD-----TQNT-----	1553
QY	1192	AEYYGLGGSSSIPTTEEPEAEPIADTVGVSVFIEVDDEMLDGLPKTDPTITSEEIMGA	1251
Db	1554	-----PAPVPPAE-----DGIK-----	1565
QY	1252	NNQARVEIAQHYNQMKCLLDENARESVAQYVNNQPPSTKVNESFRALKSINGNI-NTILS	1310
Db	1566	-----IEN-----SLKEESIEGEKEVKSTAPETAIE	1593
QY	1311	ITSQSKSHEDDT-----KPDLLNNVEMKDTAE---ETKPLRGVVDLNVVEGENIA	1359
Db	1594	CTQAPAPASEDEKVVVEPPEGEKEVKAEVKERTEEPTEPTEP--KGAADVEKVE-EKSAI	1650
QY	1360	EASGSDVVKMEEAKEEEKPKNMVV	1383
Db	1651	DLTPIVVEDKEEKEEKEEVEVML	1674

RESULT 8	
AAE36107	
ID	AAE36107 standard; protein; 1944 AA.
XX	
AC	AAE36107;
XX	
DT	26-JUN-2003 (first entry)
XX	
DE	Human chromodomain helicase DNA binding protein (CHD) #5.
XX	
KW	Human; p53 pathway; therapeutic; angiogenic disorder; apoptotic disorder;
KW	chromodomain helicase DNA binding protein; CHD; cancer; gene therapy;
KW	cell proliferative disorder; chromatin organisation modifier domain;
KW	cytostatic.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	Domain
FT	Location/Qualifiers
FT	628. .670
FT	/note= "Chromodomain"
FT	739. .1035
FT	/note= "SNF2N domain"
FT	1100. .1174
FT	/note= "Conserved C-terminal domain"



QY 1286 PP-STKVNESFRALKS 1300  
|| : : : || : : |  
Db 1856 PPIAARLQMSERSILS 1871

RESULT 9  
AAM41074  
ID AAM41074 standard; protein; 1936 AA.  
XX  
AC AAM41074;  
XX

DT 22-OCT-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 6005.  
XX

KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia.

XX Homo sapiens.

OS  
XX WO200153312-A1.

PN  
XX  
XX  
PD 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US034263.

XX 23-DEC-1999; 99US-00471275.

PR 21-JAN-2000; 2000US-00488725.

PR 25-APR-2000; 2000US-00552317.

PR 20-JUN-2000; 2000US-00598042.

PR 19-JUL-2000; 2000US-00620312.

PR 03-AUG-2000; 2000US-00653450.

PR 14-SEP-2000; 2000US-00662191.

PR 19-OCT-2000; 2000US-00693036.

PR 29-NOV-2000; 2000US-00727344.

XX (HYSE-) HYSEQ INC.

PA Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;

PI Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

DR N-PSDB; AAI60230.

XX Novel nucleic acids and polypeptides, useful for treating disorders such

PT as central nervous system injuries.

XX Example 2; SEQ ID NO 6005; 10078pp; English.

PS The invention relates to human nucleic acids (AAI57798-AAI61369) and the

XX encoded polypeptides (AAM38642-AA42213) with nootropic,

CC immunosuppressant and cytostatic activity. The polynucleotides are useful

CC in gene therapy. A composition containing a polypeptide or polynucleotide

CC of the invention may be used to treat diseases of the peripheral nervous

CC system, such as peripheral nervous injuries, peripheral neuropathy and

CC localised neuropathies and central nervous system diseases, such as

CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

CC utilisation of the activities such as: immune system suppression,

CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,

CC assays for receptor activity, arthritis and inflammation, leukaemias and

CC C.N.S disorders. Note: The sequence data for this patent did not form

XX part of the printed specification

XX Sequence 1936 AA;

Query Match		28.8%;	Score 2067.5;	DB 4;	Length 1936;
Best Local Similarity		35.3%;	Pred. No. 2.7e-147;		
Matches 517;		Conservative 205;	Mismatches 429;	Indels 313;	Gaps 44;
Qy	21	DSDDDDFVPPKDRTEFEQVEAIVRTDAKENACQACGESTNLVSCNTCTYAFHAKCLVPPL	80		
Db	447	DNSEGEIILEEVGGDLEED-----DHHMEFCRVCKDGGELLCCDTCPSSYHIHCLNPPL	501		
Qy	81	KDASVENWRCPECVSPL--NEIDKILDCE--MRPTKSSSEQSSDAE-----PKPIF---V	128		
Db	502	PEIPNGEWLCPRCTCPALKGVQKILWKWGQPPSPTPVPRPPDADPNTSPKPLEGRPE	561		
Qy	129	KQYLVKWKGLSYLHCSWVPEKB-----FQKAYKSNHRLKTRVNNFHRQME-----	173		
Db	562	RQFFVKQGMYSYWHCSWVSELQLELHCQVMFRNYQRKNMDPEPPSGDFGGDEKSRKRKN	621		
Qy	174	---SFNNSEDDF--VAIRPEWTVVDRIL-ACREEDGELEYLVKYKELSYDECYWESE---	224		
Db	622	KDPKFAEMEERFYRYGKPEWMMIHRILNHSVDKKGHVHYLIKWRDLPYDQASWESEDE	681		
Qy	225	-SDISTFQNEIQRFKDV-----NSRTRRSKQVDHKNR-----DFQQFDHTPEFLK	270		
Db	682	IQDYDLFKQSYWNHRELMRGEGRPCKLKKVLRKLERPETPTVPTVKYERQPEYLD	741		
Qy	271	---GLLHPYQLEGLNFLRFSWSKQTHVILADEMGLGKTIQSIALLASLFE--NLIPHV	325		
Db	742	ATGGTLHPYQMEGLNWLRFSAQGTDTILADEMGLGKTVQTAVFLYSLYKEGHSKGPFLV	801		
Qy	326	IAPLSTLRNWEREFATWAPQNVVMYFGTAQARAVIREHFFYLSKDQKIKKKSGQISS	385		
Db	802	SAPLSTIINWEREFEMWAPDMYVVTVYVGDKDSRAIIRENEFSFEDNAIRGGKASRM---	858		
Qy	386	ESQKRIKFDVLLTSYEMINLDSAVLKPIKWECEMIVDEGHLKKNKDSKLFSSLTQYSSNH	445		
Db	859	-KKEASVKFHVLLTSYELITIDMAILGSDWACLIVDEAHLKNNQSKFFRVNLNGYSLOH	917		
Qy	446	RILLTGTPLQNNLDELFLMHFLDAGKFGSLEEFQEEFKDINQEEQISRLHKMLAPHLLR	505		
Db	918	KLLLTGTPLQNNLEELFHLNFLTPERFHNLEGFLEEFADIAKEDIKKLHMLGPHMLR	977		
Qy	506	RVKCDVMKMPKPKELILRVDLSSLQKEYYKAIFTRNY-QVLTCKGGA-QISLNNIMMEL	563		
Db	978	RLKADVFQMPKSKTELVIVRVELSPMQKKYKYILHLSKFLKALNARGGNQVSLNVVMDL	1037		
Qy	564	RKVCCHPYMLEGVEPVJHDANEAFKQ-----LLESCGKLQLLDKMMVKLKEQGHV	614		
Db	1038	KKCCNHPYLF----PVA--AMEAPKMPNGMYDGSALIRASCKLLLLQKMLKNLKEGGHRV	1091		
Qy	615	LIYTFQFQHMLDLLEDYCTHKKQYERIDGKVGGAERQIRIDRFNAKNSNKFCLLSTRAG	674		
Db	1092	LIFSQMTKMLDLEDFLEHEGYKYERIDGGITGNMRQEAIDRFNAPGAQQFCFLSTRAG	1151		
Qy	675	GLGINLATADTVIYDSWNPHADLQAMARAHRLGQTNKVMYRLINRGTIIEERMOLTK	734		
Db	1152	GLGINLATADTVIYDSWNPNDIQAFSRAHRIQGNKKVMYRFTVTRASVEERITQVAK	1211		
Qy	735	KKMVLHLVVGK--LKTQINQEEELDDIIRYGSKELFASDEDEAGKSGK-----IHY	784		
Db	1212	KXWMLTHLVVRPGLSGTSGMSKQELDDILKFGTEELFKDEATDGGDNKEGEDSSVHY	1271		
Qy	785	DDAAIDKLLDRDLVEAEVSVDDDEENGFLKAFKVFANFEYIDENEAALAEQRAAESKS	844		
Db	1272	DDKAIERLLDRNQDETEDEL--QGMNEYLSLSSFKVAQYVVVREEMGEEEEVEEIIKQEE	1329		
Qy	845	SAGNSDRASYWEELLKDKFELHQAEEELNALGKRKRKQL-----VSIEDDLAGLEDVSS	900		
Db	1330	SVD----PDYWEKLLRHHYEQQEDLARNLGKGRIRKQVNYNDGSQEDRD---WQDDQS	1382		
Qy	901	DGDESYAEASTDG-EAAGQGVQTRRPYR---RKGRDNLEPTPLMEGEGRSFRVLGFNQS	956		
Db	1383	DNQSDYSVASEEGDEDFDERSEAPRRPSRKGLRNDKDKPLP-PLLARVGGNIEVLGFNAR	1441		
Qy	957	QRAIFVQTLMRYG-----AGNFDWKEFVPRLKQKTFEINEYGILFLKHIAEEIDENSPT	1011		



Db 1442 QRKAFLNAIMRYGMPQDAFTQW--LVRDLRGKSEKEFKAYVSLFMRHLCEPGADGAET 1499  
QY 1012 FSDGVPKEGRLIEDVLVRIALLILVQEKVKFVEDHPGKVPFSPRIILERFPGLRSGKIWKE 1071  
Db 1500 FADGVPREGLSRQHVLTRIGVMSLRKKVQEFEE----- 1532  
QY 1072 EHDKIMIRAVLKHGYGRWQAIVDVKELGIQELICKELNFPFHISLSAABQAGLQONGSGG 1131  
Db 1533 -----HVNGRW-----SMP--ELAEVEENKKMSQPGS-- 1557  
QY 1132 SNPGAQTNQPGSVITGNNNASADGAQVNSMFYYRDMQRRLLVEFVKRVLLEKAMNYYEY 1191  
Db 1558 PSPKTPSTPGD-----TQPNT----- 1575  
QY 1192 AEYYGLGSSSIPTEEPEAEPKIADTVGVSFIEVDDEMLDGLPKTDPITSEEIMGAAYD 1251  
Db 1576 -----PAPVPPAE-----DGIK----- 1587  
QY 1252 NNQARVEIAQHYNQMCKLLDENARESVOAYVNNQPPSTKVNESFRALKSINGNI-NTILS 1310  
Db 1588 -----IEEN-----SLKEEESIEGEKEVKSTAPETAIE 1615  
QY 1311 ITSQSKSHEDDT-----KPDLLNNVEMKDTAE---ETKPLRGVGVVDLNVVEGEENIA 1359  
Db 1616 CTQAPAPASEDEKVVVEPPEGEKEKVEAEVKERTPEETEPKKGGAADVEKVE-EKSAI 1674  
QY 1360 EASGSVDVVKMEAEKBEKPKNMVV 1383  
Db 1675 DLTPIVVEDKEEKKEEKEKVEML 1698

RESULT 10  
ID ABP64883  
XX ABP64883 standard; protein; 1470 AA.  
AC ABP64883;  
XX

DT 25-FEB-2003 (first entry)  
DE Human protein SEQ ID 543.  
XX

KW Human; expressed sequence tag; EST; haematopoietic disorder;  
KW central nervous system disease; viral infection;  
KW peripheral nervous system disease; non-healing wound; infectious disease;  
KW immune deficiency; immune disorder; bacterial infection; allergy; cancer;  
KW fungal infection; autoimmune disorder; coagulation disorder; neutropenic;  
KW antiallergic; antiinflammatory; immunosuppressive; neuroprotective;  
KW cytostatic; haemostatic; virucide; antibacterial; fungicide;  
KW immunostimulant; cerebroprotective.

XX Homo sapiens.  
OS  
XX WO200259260-A2.  
PN  
XX 01-AUG-2002.  
PD  
XX 16-NOV-2001; 2001WO-US042950.  
PF  
XX 17-NOV-2000; 2000US-00714936.  
PR  
XX (HYSE-) HYSEQ INC.  
PA

XX Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA;  
PI Ren F, Xue AJ, Yang Y, Wehrman T, Drmanac RT;  
PI  
XX WPI; 2002-590824/63.  
DR N-PSDB; ABQ99469.

XX New isolated polynucleotide, useful in research, diagnostic or  
PT therapeutic methods, e.g. preventing or treating disorders involving  
PT aberrant protein expression or biological activity.

XX

PS Claim 20; SEQ ID NO 543; 394pp; English.

XX The present invention relates to novel human coding sequences (ABQ99268-  
CC ABQ99608) and proteins (ABP64682-ABP65022). The sequences are useful in  
CC therapeutic, diagnostic and research methods. The polynucleotides may be  
CC used in the field of molecular biology as hybridisation probes, primers  
CC for PCR, for chromosome and gene mapping, for the recombinant production  
CC of protein, or in generation of anti-sense DNA or RNA. The  
CC polynucleotides are useful in diagnostics as expressed sequence tags  
CC (ESTs) for identifying expressed genes or for physical mapping of the  
CC human genome. The proteins may be used as molecular weight markers, or as  
CC nutritional sources or supplements. The proteins may be used to maintain  
CC and expand cell population in a totipotent or pluripotent state  
CC useful for re-engineering damaged or diseased tissues, transplantation,  
CC manufacture of bio-pharmaceuticals or the development of bio-sensors. The  
CC polynucleotides and proteins are useful for preventing, treating or  
CC ameliorating disorders involving aberrant protein expression or  
CC biological activity, e.g. haematopoietic disorders, central/peripheral  
CC nervous system diseases, mechanical and traumatic disorders, non-healing  
CC wounds, immune deficiencies and disorders, infectious diseases caused by  
CC viral, bacterial or fungal infection, autoimmune disorders, allergic  
CC reactions and conditions, coagulation disorders, or cancer. The  
CC polynucleotide sequences of the invention were assembled from ESTs  
CC isolated mainly by sequencing by hybridisation, and in some cases,  
CC sequences obtained from one or more public databases. Note: The sequence  
CC data for this patent did not form part of the printed specification, but  
CC was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX  
SQ Sequence 1470 AA;

Query Match 27.9%; Score 2007; DB 5; Length 1470;  
Best Local Similarity 36.6%; Pred. No. 7.1e-143;  
Matches 503; Conservative 212; Mismatches 387; Indels 272; Gaps 43;

QY 123 PKP---IFVKQYLKWKGLSYLHCSWVPEKEFQKAYKSNHRLKTRVNNFHR----- 170  
Db 16 PKPLEGIPEREFFVKWAGLSYWHCSWVKELQLELYHTVMYRNQKNDMDPPPPDYGSG 75  
QY 171 ----QMESFNNS-----EDDF--VAIRPEWTTVDRIIL-ACREEDGELEYLVKYKELSY 216  
Db 76 DEDGKSEKRNKDPLYAKMEERFVYGIKPEWMMIHRILNHSDFKGDVHYLIKWKOLPY 135  
QY 217 DECYWE-SESDISTFQNEIQRF-----KDVNSRTRRSKDVVDHKNRPRDFQ 260  
Db 136 DQCTWEIDDDIPYDNLKQAYWGHRELMLGEDTRLPKRLLKGGKLLRD-DKQEKPPDTP 194  
QY 261 QFDHTPEFLK-----GLLHPYQLEGLNFLRFSWSKQTHVLADENGLGKTIQSIAL 311  
Db 195 IVDPVTVKFDKQPWYIDSTGGTLHPYQLEGLNWLRFSAQQTDLADENGLGKTVQTVIF 254  
QY 312 LASLFEE--NLIPHLVIAPLSTLRNWEREFATWAPQNMVYFGTAQARAVIREHEFYLS 369  
Db 255 LYSLYKEGHSKGPYLVSAPLSTIINWEREFEMWAPDFYVVVYTYTGDKESRSVIRENEF--S 312  
QY 370 KDQKKIKKKSGQISSESKQKRIKFDVLLTSYEMINLDSAVLKPWKWECMIVDEGHLKN 429  
Db 313 FEDNAIRSGK--KVFRMKKEVQIKFHVLLTSYELITIDQAILGSIEWACLWVDEAHLKN 370  
QY 430 KDSKLFSSLTQYSSNHRILLTGTPQLQNNLDELMLMHFLDAGKFGSLEEFQEEFKDINQE 489  
Db 371 NQSKFFRVLSYKIDYKLLLTGTPLQNNLEELFHLNFLTPTPERFNNLEGFLEEFADISKE 430  
QY 490 EQISRHLKMLAPHLLRRVKDVMKDMPPKELLRLVDLSSLOKEYKKAIFTRNYQVLTKK 549  
Db 431 DQIKKLHDLGPHMLRLKADVFKNMFAKTELIVRVELSQMKQKYKFKILTRNFEALNSK 490  
QY 550 GGA-QISLNNIMMELRKVCCHPYM--LEGVE-PVIHDANEAFKQLLESCKLQLLDKMMV 605  
Db 491 GGGNQVSLNIMMDLKKCCNHPYLFPPVAAVEAPVLPNGSYDGSLSVKSGLMLLQKMLK 550  
QY 606 KLKEQGHRLVITYTFQHMULDLEDYCTHKKWQYERIDGKVGGAERQIRIDRFNAKNSNKF 665

Db 551 KLRDEGHRVLIIFSQMTKMLDLEDFLEYEGYKYERIDGGITGGLRQEAIDRFNAPGAQQF 610

Qy 666 CFLSTRAGGLGINLATADTVIIYDSWNPHADLOAMARAHRLGQTNKVMYIRLINRGTI 725

Db 611 CFLSTRAGGLGINLATADTVIIYDSWNPHNDIQAFSRAHRIGQNKVMYIRFVTRASV 670

Qy 726 EERMMQLTKKMMVLEHLVWGK---LKTQINQEEELDDIIRYGSKELFASE----- 772

Db 671 EERITQVAKRKMMTLHLVVRPGLSGSKGSMTKQELDDILKFGTELFKDDVEGMSQQR 730

Qy 773 -----DDEAGKSGK-----IHYDDAAIDKLLDRDLVEAE 801

Db 731 PVTPIPDVSSKGGNLAASAKKKHGSTPPGDNKDVEDSSVIHYDDAAISKLLDRN----- 785

Qy 802 EVSVDDEE---ENGFLKAFKVANFEYIDENEAAALEAQRVAABSKSAGNSDRASYWEEL 858

Db 786 QDATDDTELQNMNEYLSSFKAQYVVRBEDGVVEEREIIKQEE-----NVD-PDYWEKL 839

Qy 859 LKOKFELHQAELNALGKRKRSRKQL----VSI EEDDLAGLEDVSSDGDESY----EAES 910

Db 840 LRHHYEQOQEDLARNLGKGRIRKQVNYNDASQEDQE---WQBELSDNQSEYSIGSEDED 896

Qy 911 TDGEAAGQGVQTRRPYRK---GRDNL EPTPLMEGEGRSFRVLGFNQSQRAIFVQTLMR 967

Db 897 EDFEERPEG-QSGRRQSRQLKSDRDKPLP-PLLARVGGNIEVLGFNARQKAFLNAMR 954

Qy 968 YG-----AGNFDWKEFVPRLLKQKTFEINEYGILFLKHIAEEDENSPTFSFGVPKEGLR 1022

Db 955 WGMPPQDAFNHW--LVRDLRGKSEKFRAYVSLFMRHLCEPGADGAETFADGVVPREGLS 1012

Qy 1023 IEDVLVRIALLILVQEKVFEVDHPGKVPFPSPRILRFPGLRSKIKWKEEHDKIMIRAVL 1082

Db 1013 RQHVLTRIGVMSLVRKKVQEFHVNGKYSTPDLIPEGPEGKKPGEV----- 1058

Qy 1083 KHGYGRMQAIVDDKELGI---QELICKELNFPHISLSAAEQAGLQGGNGSGSNPGAQT 1138

Db 1059 -----ISSDPNTPVPASPAPLHPAPLGLPD---KMEAQLGYMDE----KDPGAQK 1101

Qy 1139 NONPGSVITGNNNASADGAQVNSMFYYRDMQRRLVFVKRVLLEKAMNVEYAEYYGL 1198

Db 1102 PRQP-----LEVOALPAALDRVESEDKH-- 1124

Qy 1199 GGSSSIPTPEPEAEPKIADTVGVSFIEVDDEMLDGLPKTDPITSEHIMGAADVNNQARVE 1258

Db 1125 -----ESPASKERARE-----ERPEETEKAPPSPEQLPREVL----- 1157

Qy 1259 IAQHYNQMKLLDENARESVOAYVNN---OPPSTKVNESFRALKSINGNINTILSITSDQ 1315

Db 1158 -----PEKEKILDKLELSLIHSRGSSELRPDDTKAEKEPIETQQNGD----- 1201

Qy 1316 SKSHEDTKPDNLNNVEMK---DTAETKPLRGGVVDNLNV-EGEENIAEASGSV 1365

Db 1202 -KEEDEGKEDKKGKFKFMFIAD-----GGFTLHTLWQNEERAASVSGKI 1248

RESULT 11

AAE36108

ID AAE36108 standard; protein; 1377 AA.

XX

AC AAE36108;

XX

DT 26-JUN-2003 (first entry)

XX

DE Human chromodomain helicase DNA binding protein (CHD) #6.

XX

KW Human; p53 pathway; therapeutic; angiogenic disorder; apoptotic disorder;

KW chromodomain helicase DNA binding protein; CHD; cancer; gene therapy;

KW cell proliferative disorder; chromatin organisation modifier domain;

KW cytosstatic.

XX

OS Homo sapiens.

XX

FH Location/Qualifiers

Domain 12. .54

FT /note= "Chromodomain"

FT 126. .422

FT /note= "SNF2N domain"

FT 487. .561

FT /note= "Conserved C-terminal domain"

XX

WO200298899-A2.

PN

XX

PD 12-DEC-2002.

XX

PF 03-JUN-2002; 2002WO-US017466.

XX

PR 05-JUN-2001; 2001US-0296076P.

PR 10-OCT-2001; 2001US-0328605P.

PR 22-OCT-2001; 2001US-0338733P.

PR 15-FEB-2002; 2002US-0357253P.

PR 15-FEB-2002; 2002US-0357600P.

XX

(EXEL-) EXELIXIS INC.

XX

PI Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;

PI Lioubin MN;

XX

DR WPI; 2003-156840/15.

XX

Identifying a candidate p53 pathway-modulating agent as therapeutic targets for disorders related to defective p53 function e.g. cancer by contacting an assay system having purified CHD polypeptide or nucleic acid, with a test agent.

XX

Claim 13; Page 215-221; 278pp; English.

XX

The present invention relates to a method for identifying candidate p53 pathway modulating agents. The method involves contacting an assay system comprising purified chromatin organisation modifier (chromo) domain helicase DNA binding proteins (CHD), nucleic acids, their functionally active fragments or derivatives, with a test agent under conditions where, but for the presence of the test agent, the system provides a reference activity. The methods are useful for identifying modulators of the p53 pathway as therapeutic targets for disorders associated with defective p53 function, such as angiogenic disorders, apoptotic disorders or cell proliferative disorders, e.g. cancer. The modulators are useful as research reagents, diagnostics and therapeutics. The invention is also useful in gene therapy. The present sequence is human chromodomain helicase DNA binding protein

XX

SQ Sequence 1377 AA;

Query Match 27.0%; Score 1939.5; DB 6; Length 1377;

Best Local Similarity 37.3%; Pred. No. 8.8e-138;

Matches 480; Conservative 202; Mismatches 358; Indels 247; Gaps 39;

Qy 185 IRPEWTTVDRII-ACREEDGELEYLVKYLKELSYDECYWE-SESDISTFQNEIQRF----- 237

Db 10 IKPEWMMIHRILNHSFDKGDVHYLIKWKDLDPYDQCTWEIDDDIDIPYDNLKQAYWGHRE 69

Qy 238 -----KDVNSRTRRSKDVHDHKNRPDRFQQFDHTPEFLK-----GLLHPYQL 278

Db 70 LMLGEDTRLPKRLKKGKLRD-DKQEKPPDTPIDPTVKFDKQPWYIDSTGGTLHPYQL 128

Qy 279 EGLNFLRFSWSKQTHVILADEMGLGKTIQSIALLASLFEE--NLIPHVLVIAPLSTLRNWE 336

Db 129 EGLNWLRFWSAQGTDITILADEMGLGKTVQTIIVFLYSLYKBSKGPYLVSAPLSTIINWE 188

Qy 337 REFATWAPQNMVVMYFGTAQARAVIREHEFYLSKDQKIKKKSGQISSSESKQRIKFDV 396

Db 189 REFEMWAPDFYVVVYTTGDKESRSVIRENEF--SFEDNAIRSGK--KVFRMKKEVQIKFHV 244

Qy 397 LLTSYEMINLDSAVLKPPIKWECEMIVDEGHRLLKNKDSKLFSSLTQYSSNHRILLTGTPLQN 456

Db 245 LLTSYELITIDQAILGSIWACLIVVDEAHRLLKNNQSKFFRVLNSYKIDYKLLLTGTPLQN 304





Qy 171 ----QMESFNNS-----EDDF--VAIRPEWTTVDRI--ACREEDGELEYLVKYKELSY 216  
Db 76 DEDGKSEKRNKDPYAKMEERFYRYGIKPEWMMIHRILNHSFDKGDVHYLIKWKOLPY 135  
Qy 217 DECYWE-SESDISTFQNEIQRF-----KDVNSRTRRSKDVHDKRNPRDFQ 260  
Db 136 DQCTWEIDIDIPYDNLKQAYWGHRELMLGEDTRLPKRLKKGKKLRD-DKREKPPDTP 194  
Qy 261 QFDHTPEFLK-----GLLHPYQLEGLNFLRFSWSKQTHVILADEMGLKTIQSIALL 311  
Db 195 IVDPTVKFDQWPYIDSTGGTLHPYQLEGLNWLRFWSAQGTDTILADEMGLKTVQTIIVF 254  
Qy 312 LASLFEE--NLIPHLVIAPLTLRNWEREFATWAPQNNVVMYFGTAQARAVIREHEFYLS 369  
Db 255 LYSLYKEGHSKGPYLVSAPLSTIINWEREFEMWAPDFYVVVYTGDKESRSVIRENEFSFG 314  
Qy 370 KOQKKIKKKSGQISSESKQRIKFDVLLTSYEMINLDSAVLKPIKWECMIVDEGHLRKN 429  
Db 315 DNAIRSGKK---VFRMKKEVQIKFHVLLTSYELITIDQAILGSIWACLWVDEAHLRKN 370  
Qy 430 KDSKLFSSLTQYSSNHRILLTGTPQNNLDELFMHFLDAGKFGSLSEEFQEEFKOINQE 489  
Db 371 NQSKFFRVLSYKIDYKLLLTGTPLQNNLELFLHLLNFLTPERFNNLEGFLEEFADISKE 430  
Qy 490 EQISRLHKLAPHLLRRVKDKVMKMPKKEILRLVDLSSLOKEYKAIFTRNYQVLTKK 549  
Db 431 DQIKRLHDLGPHMLRRLKADVFKMMPAKTELIIVRVELSQMKKYKFILTRNFEALNSK 490  
Qy 550 GGA-QISLNMIMMELRKVCCHPYM--LEGVE-PVIHDANEAFKQLLES CGKLQLLDKMMV 605  
Db 491 GGGNQVSLNIMMDLKKCCNHPYLPFVAAPVLPNGSNDGSSLVKSSGKMLLQKMLK 550  
Qy 606 KLKEQHRVLIYTFQHMLDLLEDYCTHKKQYERIDGKVGABEQIRIDRFNAKNSKF 665  
Db 551 KLRDEGHRVLIFSQMTKMLDLEDLFLEYGYKYERIDGGITGGLRQEAIDRFNAPGAQQF 610  
Qy 666 CFLSTRAGLGINLATADTVIIYDSWNPHADLQAMARAHRLGQTNKVMYRLINRGTI 725  
Db 611 CFLSTRASGLGINLATADTVIIYDSWNPHNDIQAFSRAHRIQGNKKVMYRFVTRASV 670  
Qy 726 EERMQLTKKMWLEHLVVGK--LKTQINQEEELDDIIRYGSKELFASE----- 772  
Db 671 EERITQVAKRKQMLTHLVVRPGLSGSGSMTKQELDDLKFPGTTELKDDVEGMMSQQR 730  
Qy 773 -----DDEAGKSGK-----IHVDDAAIDKLLDRDLVEAE 801  
Db 731 PVTPIPDVQSSKGGNLAASAKKHGHTPPGDNKDVEDSSVIHYDDAAISKLLDRN----- 785  
Qy 802 EVSVDDEE---ENGFLKAFKQVANFEYIDENEAALAEQORVAESKSSAGNSDRASYWEEL 858  
Db 786 QDATDDTELQNMNEYLSSFKVAQYVVREEDGVEGVEREIIKQEE----NVD-PDYWEKL 839  
Qy 859 LKDKFELHQAEBELNALGKRKRQKQL---VSI EDDDLAGLEDVSSDGDESY---EAE 910  
Db 840 LRHHEQQQEDLARNLGKGRIRKQVNYNDASQEDQE---WQDEISDNQSEYSIGSEDED 896  
Qy 911 TDGEAAGQGVOTGRRPYRRK---GRDNL EPTPLMEGEGRSFRVLGPNQSQRAIFVQTLMR 967  
Db 897 EDFEERPEG-QSGRRQSRQLKSDRDKPLP-PLLARVGGNIEVLGPNARQKAFLNAMR 954  
Qy 968 YG-----AGNFDW 975  
Db 955 WGMPPQDAFNHSHW 967

RESULT 13  
ABB71771  
ID ABB71771 standard; protein; 883 AA.  
XX  
AC ABB71771;  
XX  
DT 26-MAR-2002 (first entry)  
XX

DE Drosophila melanogaster polypeptide SEQ ID NO 42105.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.  
XX Drosophila melanogaster.  
OS  
XX WO200171042-A2.  
PN  
XX 27-SEP-2001.  
PD  
XX 23-MAR-2001; 2001WO-US009231.  
PF  
XX 23-MAR-2000; 2000US-0191637P.  
PR  
XX 11-JUL-2000; 2000US-00614150.  
PR  
XX (PEKE ) PE CORP NY.  
PA  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI  
XX WPI; 2001-656860/75.  
DR N-PSDB; ABL15874.  
DR  
XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signaling and cell-cell interactions.  
PT  
PT  
PT  
XX Disclosure; SEQ ID NO 42105; 21pp + Sequence Listing; English.  
PS  
XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 883 AA;  
Query Match 24.2%; Score 1738.5; DB 4; Length 883;  
Best Local Similarity 42.5%; Pred. No. 8.8e-123;  
Matches 377; Conservative 152; Mismatches 256; Indels 103; Gaps 23;  
Qy 44 RTDAK-----ENACQACGESTNLVSCNTCTYAFHAKLVPPPLKDASVENWRCECVSPL 97  
Db 24 KTNAKKQKFRDEEYKVCSDGGDLLCCDSCPSVYHRTCLSPPLKSIKPGDWICPRCIPLP 83  
Qy 98 NEIDKILDCEMRPTKSSEQSSDAEPKPIFVKQYLVKWKGLSYLHCSWVPEKEFQKAYS 157  
Db 84 GKAEKILSWRWALDRSVELRTSKGEK---REYFIKWHGMSYWHCEWIPEGQMLLHAS 139  
Qy 158 NHRLKTRVNNFHRQ-----MESFNNS EDDF-----VAIRPEWTTVDRI LACREE-DG 203  
Db 140 -----MVASFQRRSDMEEPSLEELDQDGNLHERFYRYGIKPPEWLLVQRVINHSEEPNG 193  
Qy 204 ELEYLVKYKELSYDECYWESESD-ISTFQNEIQRFKDVNS-----RTRRSKDVHDKRNP 256  
Db 194 GTMYLVKWRELSYNDSSWERESDSIPGLNQAIALYKILRSNKGQRDRPAPTIDLNK-- 251  
Qy 257 RDPQOPDHTPEFLK--GL-LHPYQLEGLNFLRFSWSKQTHVILADEMGLKTIQSIALLA 313  
Db 252 ----KYEDQPVFLKEAGLKLHPFQIEGVSWLRYSWGQGTILADEMGLKTIQTVVFLY 307  
Qy 314 SLFEENLI--PHLVIAPLTLRNWEREFATWAPQNNVVMYFGTAQARAVIREHEFYLSKD 371  
Db 308 SLPKBGHCRGPFLLISVPLSTLNWERELELWAPELYCVTVGGKTARAVIRKHEISFEV 367  
Qy 372 QKIKKKKSGQISSESKQRIKFDVLLTSYEMINLDSAVLKPIKWECMIVDEGHLRKNKD 431  
Db 368 TTKTMRE-----NQYKFNVMLTSEYFISVDAFLGCIDWAALVDEAHLRSNQ 418



Db 856 DGESEDFCELLSTRAGGLGINLASADTVVIFSDWNPQNDLQARAHRIQKKQVNIYRL 915  
QY 720 INRGTIERNMQLTKKMVLHLV-----GKLTQINQEEELDDIIRYG 764  
Db 916 VTKGVIEEIIERAKKQMVLDHLVIQRMDDTGTRTILENNSGRNSNPFNKEELTAILKFG 975  
QY 765 SKELFASDEDEAGSKGIHYDDAAIDKLLDRDLVEAEVSVDDDEENGFLKAFKVANFEY 824  
Db 976 AEDLFKELEGESEP-----QEMDIDEIL--RLAETRENEVSTSATDELLSQFKVANFAT 1028  
QY 825 IDENAAAALAEQRAVAEASKSSAGNSDRASYWEELL-----KDKFELHQAEEELNALGK 876  
Db 1029 MEDEE--ELEER-----PHKD---WDEIIPPEQRKKVVEEERQKELEEIYMLPR 1072  
QY 877 RKRSRKQLVSIEDDLAGLEDVSSDGDSEYAE--STDGEAAGQGVQGTGRRPYRRKGRDNL 935  
Db 1073 IRSSTKK-----AQTNDSDSTESKRAQORSSASESETEDSDDDDKPKRR----- 1117  
QY 936 EPTPLMEGEGRSFR---VLGFNQSORAIFVQTLRMRYGAGNFDWKEFVPR---LKQKTFEE 989  
Db 1118 -----GRPRSVRKDLVEGFTDAEIRRFIKAYKKFGL-PLERLECLARDAELVDKSVAD 1169  
QY 990 INEYGILF-----LKHIAEEIDENSPTFSDBGVPKEGLRIEDVLVRIALLILVQEKVKF 1042  
Db 1170 LKRLGELTHNSCVSAMQYEEQLENASEGKPGKRRGPTIKISGVQVNVKSIQHEEF 1229  
QY 1043 VEDHPGKVPFPPSR-----ILERFPGLSGKIWKEEHDKIMIRAVLKHGGRWQAIIVDDKE 1097  
Db 1230 EMLHKSIPVDPEKKKYCLTCRVKAAHFDVWGVEDDSRLLLGIYEHGYGNWELIKTDPE 1289  
QY 1098 LGIQELIC-----KELNFPHISLSAAEQAGLQONGSGSNPGAQTNQNPGSVIT 1147  
Db 1290 LKLTDKILPVETDKKPOGKQLQTRADYLLKLRKGLEK-----GAVT 1332  
QY 1148 GNNASADGAQVNSMFYRRDMQRRVLVEFVKRVLLLEKAMNVEYAEYGLGGSSSIPT 1207  
Db 1333 GGEAK-----LKKRKPRVKKENKVPRLKEEHGIELSSPRHSD 1370  
QY 1208 EPEAEPKIADTVGSFIEVDDEMLDGLPKTDPITSEEIMGAADVNNQARVEIAQHYNQMC 1267  
Db 1371 NPSEEGEVKD-----DGLEKS-PMKKQKKKENKENKE-----KQMS 1406  
QY 1268 KLLD-ENARESQAQVNNQPPSTKVNESFRALKSINGNINTILSITSDQSKSHEDTKPD 1326  
Db 1407 SRDKKEGDKERKSKDKKKEPKSGDAKSSKSRSQGPVHITAGSEPVPIGEDEDD---D 1463  
QY 1327 LNNVEMKDTAEETKPLRGGVVLDLNVVEGEENIAE 1360  
Db 1464 LDQETFSICKERMVRPVKKALKQOLDKPDKGLNVQE 1497

RESULT 15  
AAE36106  
ID AAE36106 standard; protein; 1739 AA.  
XX  
AC AAE36106;  
XX  
DT 26-JUN-2003 (first entry)  
XX  
DE Human chromodomain helicase DNA binding protein (CHD) #4.  
XX  
KW Human; p53 pathway; therapeutic; angiogenic disorder; apoptotic disorder;  
KW chromodomain helicase DNA binding protein; CHD; cancer; gene therapy;  
KW cell proliferative disorder; chromatin organisation modifier domain;  
KW cytosstatic.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Domain 281..327  
FT /note= "Chromodomain"  
FT Domain 375..433

FT Domain /note= "Chromodomain"  
FT 487..768  
FT /note= "SNF2N domain"  
FT 831..905  
FT /note= "Conserved C-terminal domain"  
XX  
PN WO200298899-A2.  
XX  
PD 12-DEC-2002.  
XX  
PF 03-JUN-2002; 2002WO-US017466.  
XX  
PR 05-JUN-2001; 2001US-0296076P.  
PR 10-OCT-2001; 2001US-0328605P.  
PR 22-OCT-2001; 2001US-0338733P.  
PR 15-FEB-2002; 2002US-0357253P.  
PR 15-FEB-2002; 2002US-0357600P.  
XX  
PA (EXEL-) EXELIXIS INC.  
XX  
PI Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;  
PI Lioubin MN;  
XX  
DR WPI; 2003-156840/15.  
XX  
PT Identifying a candidate p53 pathway-modulating agent as therapeutic  
PT targets for disorders related to defective p53 function e.g. cancer by  
PT contacting an assay system having purified CHD polypeptide or nucleic  
PT acid, with a test agent.  
XX  
PS Claim 13; Page 192-199; 278pp; English.  
XX  
CC The present invention relates to a method for identifying candidate p53  
CC pathway modulating agents. The method involves contacting an assay system  
CC comprising purified chromatin organisation modifier (chromo) domain  
CC helicase DNA binding proteins (CHD), nucleic acids, their functionally  
CC active fragments or derivatives, with a test agent under conditions  
CC where, but for the presence of the test agent, the system provides a  
CC reference activity. The methods are useful for identifying modulators of  
CC the p53 pathway as therapeutic targets for disorders associated with  
CC defective p53 function, such as angiogenic disorders, apoptotic disorders  
CC or cell proliferative disorders, e.g. cancer. The modulators are useful  
CC as research reagents, diagnostics and therapeutics. The invention is also  
CC useful in gene therapy. The present sequence is human chromodomain  
CC helicase DNA binding protein  
XX  
SQ Sequence 1739 AA;

Query Match 22.6%; Score 1621.5; DB 6; Length 1739;  
Best Local Similarity 30.7%; Pred. No. 2e-113;  
Matches 453; Conservative 229; Mismatches 493; Indels 299; Gaps 43;  
QY 11 RSDRKPVYNLDDSDDDDFVPKK-----DRTFEQVEAIVRTDAKENACQACG 56  
Db 199 RGKRKKQDSSDEDDDDDEAPKQTRRRRAAKNVSYKEDDDFE-----TDS-DDLIEMTG 250  
QY 57 ESTNLVSCNTCTYAFHAKCLVPPPLKASVENWRCPECVSPLEIDKILDCEMRPTKSSEQ 116  
Db 251 EGVDEQQDQDNSET-----IEKVLDD--SRLGKKGAT 277  
QY 117 GSS-----DAEPKPIFVK-----QYLVKWKGLSYLHCSWVPEKEFQ----KAYKS 157  
Db 278 GASTTVYAIENGDPSPGDFTEKDEGEIQYLIKWKGWSYIHTWESESLQQQVKVGLKK 337  
QY 158 NHRLKTRVNNFHR-----QMESFNNSDEDDFVAIRPEWTTVDRLACREE----- 201  
Db 338 LENFKKKEDEIKQWLGVKSPEDVEYFNCQELASELNNKQYQIVERVIAVTKSTLQOTD 397  
QY 202 -----DGELEYLVKYKELSYDECYWESEDI-STFQNEIQRFKDVNSR----TRR 246  
Db 398 FFAHSRKPAAPSNEPEYLCWKWGLPYSECSWEDEALIGKKFQNCIDSFHSRNNSKTIPTRE 457  
QY 247 SKDVDHKRNPRDFQQFDHTPEFLKG---LLHPYQLEGLNFLRFSWSKQTHVILADEMGLG 303



Db 458 CKAL--KORPR-FVALKKQPAYLGGENLELDYQLEGLNWLASHWCXNNSVILADEMGLG 514  
QY 304 KTIQSIALLASLEENLI--PHLVIAPLSTLRNWEREFATWAPQNVVMVYFGTAQARAVI 361  
Db 515 KTIQTISFLSFHQHQLYGPFLIVPLSTLTSWOREFEIWAPEINVVYIGDLMsrNTI 574  
QY 362 REHEFYLSKDQKIKKKKSGQISSESQKRIKFDVLLTSYEMINLDSAVLKPIKWECMIV 421  
Db 575 REYEWI-----HSQTKRLKFNALITTYEILLKDKTVLGSINWAF LGV 616  
QY 422 DEGHRLLKNKDSKLFSSLTQYSSNHRILLTGTPLONNLDELFLMHFLDAGKFGSLEEFQE 481  
Db 617 DEAHRLKNDSDSLLYKTLIDFKSNHRLLLITGTPLONSLKELWSLLHFLIMPEKFEFWEDFEE 676  
QY 482 EFKDINQEEQISRLHKMLAPHLLRRVKVDVMDPPKELILRVDLSSLOKEYYKAIFTR 541  
Db 677 DHGK-GRENGYQSLHKVLEPFLRLRRVKDVKS LPAKVEQILRVEMSALQKYKWLTR 735  
QY 542 NYQVLTK-KGGAQISLNNIMMELRKVCCHPYMLEGVEP-VIHDANEAFKQLLES CGKLQ 599  
Db 736 NYKALAKGTRGTSGLNIVMELKKCCNHCYLIKPPENERENGQEILLSLRSSGKLIL 795  
QY 600 LDKMMVKLKEQGHVRVLIYTOFQHMLDLLEDYCTHKKWQYERIDGKVGGAERQIRIDRFNA 659  
Db 796 LDKLLTRLRERGNRVLIIFSQVMRLDILAEYLTIKHYPPQRLDGSIKGEIRKQALDHfNA 855  
QY 660 KNSNKFCFLSTRAGGLGINLATADTVIYVSDWNPHADLOAMARAHRLGQTNKVMYRL 719  
Db 856 DGEDFCFLSTRAGGLGINLASADTVIFDSDNWPNQDLQAARAHRI GQKQVNIYRL 915  
QY 720 INRGTIERMMQLTKKKMWLEHLV-----GKLKTONINQEEELDDIIRY G 764  
Db 916 VTKGTVEEEIIERAKKQWLDHLVIQRMDDTTGRTILENNSGRSNSNPENKBEILTAILKFG 975  
QY 765 SKELFASDEDEAGSKGIHYDDAAIDKLLDRDLVEAAEVSVDDEEENGFLKAFKVANFEY 824  
Db 976 AEDLFKELEGESEP-----QEMDIDEIL--PLAETRENEVSTSATDELLSQPKVANFAT 1028  
QY 825 IDENEAALAEQORVAESKSSAGNSDRASYWELL-----KDKFELHQAEEELNALGK 876  
Db 1029 MEDEE--ELEER-----PHKD-----WDEIIPEEQRKKVEEERQKELEEIYMLPR 1072  
QY 877 RKRSRKQLVSI EEDDLAGLEDVSSDGESEYAE-STDGEAAGQVQTGRRPYRRKGRDNL 935  
Db 1073 IRSSTKK-----AQTNDSDSOTESKROQRSSASESETEDSDDDKKPKRR----- 1117  
QY 936 EPTPLMEGEGRSFR---VLGFNQSORAIFVQTLNRYGAGNFDWKEFVPR---LKQKT FEE 989  
Db 1118 -----GRPRSVRKDLVEGFTDAEIRRFIKAYKKFGL-PLERLECLARDAELVDKSVAD 1169  
QY 990 INEYGILF-----LKHIAEEIDENSPTFSDGVPKEGLRIEDVLVRIALLILVQEKVKF 1042  
Db 1170 LKRLGELIHNCSVSAQYEEYEQLENASEGKGPGRRGPTIKISGVQNVVKSIIQH EEEF 1229  
QY 1043 VEDHPGKPVFPSR-----ILERFPGLRSGKIWKEEHDKIMIRAVLKHG YGRWQAI VDDKE 1097  
Db 1230 EMLHKSIPVDPEEKKKYCLTCRVKAAHFDVEGWVEDDSRLLLGIYEHGYGNWELIKTDPE 1289  
QY 1098 LGIQELIC-----KELNFPHISLSAAEQAGLQGGSGGSPGAGTQNQNGPSVIT 1147  
Db 1290 LKLTDKILPVETDKKPOGKQLQTRADYLLKLLRKGLEKK-----GAVT 1332  
QY 1148 GNNNASADGAQVNSMFYYRDMQRRLLVEFVKRVLLLEKAMNYEAYEYGLGSSSSIPTE 1207  
Db 1333 GGEAK-----LKKRPRVKKENKVPRLKEEHGIELSSPRHSD 1370  
QY 1208 EPEAEPKIADTVGVSFIEVDDEMGLPKTDPITSEEIMGAADVNNQARVEIAQHYNQMC 1267  
Db 1371 NPSEEGEVKD-----DGLEKS-PMKKKQKKKENKENKE-----KQMS 1406  
QY 1268 KLID-ENARES VQAYVNNQPPSTKVNESFRALKSINGNINTILSITSDQSKSHEDDTKPD 1326

Db 1407 SRKDXEGDKERKKSKDKKKEPKSGDAKSSSKRSQGPVHITAGSEPVPIGEDEDD---D 1463  
QY 1327 LNNVEMKDTAEBETKPLRGGVVDLNVVEGEENIAE 1360  
Db 1464 LDQETFSICKERMMPVKKALKQLDKPDKGLNVQE 1497

Search completed: September 14, 2004, 02:04:22  
Job time : 173 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 14, 2004, 02:01:31 ; Search time 36 Seconds  
(without alignments)  
1984.732 Million cell updates/sec

Title: US-10-049-137-2  
Perfect score: 7187  
Sequence: 1 MSSSLVERLIRSDRKPVYNL.....VDVKMEEAKBEEKPKNMVVD 1384

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2091	29.1	1912	4	US-08-913-832A-2
2	2091	29.1	1912	4	US-09-249-181A-2
3	2091	29.1	1912	4	US-09-158-707-2
4	1623.5	22.6	1739	4	US-09-976-594-76
5	1028.5	14.3	1646	4	US-09-535-008-67
6	1027	14.3	1647	4	US-09-535-008-2
7	1027	14.3	1649	4	US-09-535-008-75
8	1025.5	14.3	1650	4	US-09-535-008-71
9	1011	14.1	1681	4	US-09-535-008-77
10	1009.5	14.0	1678	4	US-09-535-008-69
11	1009.5	14.0	1682	4	US-09-535-008-73
12	1008	14.0	1679	4	US-09-535-008-65
13	997	13.9	831	4	US-09-702-705-1819
14	997	13.9	831	4	US-09-736-457-1819
15	997	13.9	831	4	US-09-671-325-1819
16	898.5	12.5	801	3	US-09-104-070-2
17	699.5	9.7	904	4	US-09-252-991A-29119
18	648	9.0	1070	4	US-09-107-532A-3716
19	620	8.6	1073	4	US-09-134-000C-6356
20	619.5	8.6	1247	4	US-09-404-627-2
21	537	7.5	2972	4	US-09-579-181-2
22	537	7.5	3118	4	US-09-579-181-1
23	459	6.4	1075	4	US-09-198-452A-916
24	402.5	5.6	727	3	US-09-179-558-56
25	402.5	5.6	727	4	US-09-722-825-56
26	402.5	5.6	727	4	US-09-722-487-56
27	402.5	5.6	727	4	US-09-722-708-56

28	402.5	5.6	941	3	US-09-179-558-55	Sequence 55, Appl
29	402.5	5.6	941	4	US-09-722-825-55	Sequence 55, Appl
30	402.5	5.6	941	4	US-09-722-487-55	Sequence 55, Appl
31	402.5	5.6	941	4	US-09-722-708-55	Sequence 55, Appl
32	402	5.6	949	4	US-09-404-627-4	Sequence 4, Appli
33	331	4.6	946	4	US-09-252-991A-18989	Sequence 18989, A
34	318	4.4	722	4	US-09-252-991A-26452	Sequence 26452, A
35	306.5	4.3	459	4	US-09-134-000C-6639	Sequence 6639, Ap
36	241	3.4	301	4	US-09-198-452A-898	Sequence 898, App
37	239	3.3	359	3	US-09-179-558-65	Sequence 65, Appl
38	239	3.3	359	4	US-09-722-825-65	Sequence 65, Appl
39	239	3.3	359	4	US-09-722-487-65	Sequence 65, Appl
40	239	3.3	359	4	US-09-722-708-65	Sequence 65, Appl
41	233.5	3.2	951	4	US-09-328-352-4456	Sequence 4456, Ap
42	220.5	3.1	247	4	US-09-370-838-112	Sequence 112, App
43	220.5	3.1	1003	4	US-09-489-039A-12357	Sequence 12357, A
44	220	3.1	351	4	US-07-945-295-2	Sequence 2, Appli
45	220	3.1	351	5	PCT-US91-06418-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-08-913-832A-2  
; Sequence 2, Application US/08913832A  
; Patent No. 6329517  
; GENERAL INFORMATION:  
; APPLICANT: Seelig, Hans Peter  
; APPLICANT: Renz, Manfred  
; TITLE OF INVENTION: DERMATOMYOSITIS-SPECIFIC AUTO-ANTIGEN  
; FILE REFERENCE: 8484-0030-999  
; CURRENT APPLICATION NUMBER: US/08/913,832A  
; CURRENT FILING DATE: 1998-01-12  
; PRIOR APPLICATION NUMBER: PCT/DE96/00444  
; PRIOR FILING DATE: 1996-03-08  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 1912  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-08-913-832A-2

Query Match	29.1%	Score	2091;	DB	4;	Length	1912;
Best Local Similarity	35.5%	Pred. No.	8.3e-166;				
Matches	520;	Conservative	205;	Mismatches	424;	Indels	314;
Gaps	44;						
Qy	21	DDSDDDDFVPKKORTFEQVEAIVRTDAKENACQACGESTNLVSCNTCTYAFHAKCLVPPL	80				
Db	426	DNSEGEILEEVGGDLEED-----DHHMEFCRVCKDGGELLCDDTCPSSTYHICLNPPPL	480				
Qy	81	KDASVENWRCPECVSPL--NEIDKILDCE--MRPTKSSEQSSDAB-----PKPIF---V	128				
Db	481	PEIPNGEWLCPRCTCPALKGKVKILWKWGQPPSPTPVPRPPDADPNTSPKPLEGRPE	540				
Qy	129	KQYLVKWKGLSYLHCSWVPEKE-----FQKAYKSNHRLKTRVNNFHRQME-----	173				
Db	541	RQFFVKQGMYSYWHCSWVSELQELHCVQMFNRYQKNDMDPPSGDFGGDEKSRKRKN	600				
Qy	174	---SFNNSEDDF--VAIRPEWTVDRIL-ACREEDGELEVLVKYKELSYDECYWESE---	224				
Db	601	KDPKFAEMEERFYRYGIKPEWMMIHRILNHSVDKKGHVHYLIKWRDLFYDQASWESE	660				
Qy	225	-SDISTFQNEIQRFKDV-----NSRTRRSKQVDHKKRNP-----DFQFDHTPEFLK	270				
Db	661	IQDYDLFKQSYNNHRELMRGEGRPKKLKKVLRKLERPPETPTVDTVKYERQPEYLD	720				
Qy	271	---GULLHPYQLEGLNFRFSKQTHVILADEMGLGKTIQSIALLASLFEE--NLIPHLV	325				
Db	721	ATGGTLHPYQMEGLNWLRFSAQGTDTILADEMGLGKTVQTAFLYSLYKEGHSKGPFLV	780				
Qy	326	IAPLSTLRNWEREFATWAPQNVVMVYFGTAQARAVIREHEFVLSKDQKIKKKKSGQISS	385				







Db 1362 NQSDYSVASEGDEDFDERSEAPRRPSRKGLRNDKDKPLP-PLLARVCGNIEVLGFNARQ 1420

Qy 958 RAIFVQTLRMVG-----AGNFDWKEFVPRLKQKTFEEINEYIGILFLKHIAEIDENSPTF 1012

Db 1421 RKAFLNAIMRYGMPPQDAFTQW--LVRLRGKSEKEFKAYVSLFMRHLCPEGADGETF 1478

Qy 1013 SDGVPKEGLRIEDVLVRIALLILVQEKVKEVEDHPGKPVFPSPRILRRPGLRSGKIWKEE 1072

Db 1479 ADGVPREGLSRQHLVTRIGVMSLIRKKVQEFB----- 1510

Qy 1073 HDKIMIRAVLKHGYGRWQAIIVDDKELGIQELICKELNFPHISLSAAEQAGLQGQNGSGGS 1132

Db 1511 -----HVNGRW-----SMP--ELAEVEENKKWSPQGS--P 1536

Qy 1133 NPGAQTNQPGSVITGNNNASADGAQVNSMFYYRDMQRRLLVEFVKRVLLEKAMNYEYA 1192

Db 1537 SPKTPTSTPGD-----TQPNT----- 1553

Qy 1193 EEEYGLGGSSSIPTEEPEAEPKIADTVGVSPFIEVDDEMLDGLPKTDPTITSEEIMGAAVDN 1252

Db 1554 -----PAPVPPAE-----DGIK----- 1565

Qy 1253 NOARVEIAQHYNQMKLLDENARESVOAYVNNQPPSTKVNESFRALKSINGNI-NTILSI 1311

Db 1566 -----IEN-----SLKEEESIEGEKEVKSTAPETAIEC 1594

Qy 1312 TSDQSKSHEDT-----KPDLNNVEMKDTAE---ETKPLRGVVDLNVVEGEENIAE 1360

Db 1595 TQAPAPASEDEKVVPPEPEGEKEVKAEVKERTEEPMETEP--KGAADVEKVE-EKSAID 1651

Qy 1361 ASGSVDVKMEEAKEEEKPKNMVV 1383

Db 1652 LTPIVVEDKEEKEEKEKEVML 1674

RESULT 4

US-09-976-594-76

; Sequence 76, Application US/09976594

; Patent No. 6673549

; GENERAL INFORMATION:

; APPLICANT: Furness, Michael

; APPLICANT: Buchbinder, Jenny

; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS

; FILE REFERENCE: PA-0041 US

; CURRENT APPLICATION NUMBER: US/09/976,594

; CURRENT FILING DATE: 2001-10-12

; PRIOR APPLICATION NUMBER: 60/240,409

; PRIOR FILING DATE: 2000-10-12

; NUMBER OF SEQ ID NOS: 1143

; SOFTWARE: PERL Program

; SEQ ID NO 76

; LENGTH: 1739

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; OTHER INFORMATION: Incyte ID No. 6673549 3125723CD1

US-09-976-594-76

Query Match 22.6%; Score 1623.5; DB 4; Length 1739;

Best Local Similarity 30.7%; Pred. No. 1.2e-126;

Matches 453; Conservative 229; Mismatches 493; Indels 299; Gaps 43;

Qy 11 RSDRKPVYNLDDSDDDFVPKK-----DRTFEQVEAIVRTDAKENACQACG 56

Db 199 RGKRKKQDSSDEDDDEAPKQTRRRRAAKNVSYKEDDDFE-----TDS-DDLIEMTG 250

Qy 57 ESTNLVSCNTCTYAFHAKCLVPPLKDVASVENWRCPVCVSPNLNEIDKILDCEMRPTKSSEQ 116

Db 251 EGVDEQQDNSET-----IEKVLD--SRLGKKGAT 277

Qy 117 GSS-----DABPKPIFVK-----QYLVKWKGLSYLHCSWVPEKEFQ-----KAYKS 157

Db 278 GASITVVAIEANGDPSGDFDTEKDEGEIQYLIKWKGWSYHSTWSEESLQQQKVGLKK 337

Qy 158 NHRLKTRVNNFHR-----QMESENNSEDDFVAIRPEWTTVDRIACREE----- 201

Db 338 LENFKKKEDEIKQWLGVKSPEDVEYFNCQELASELNKQOIVERVIAVKTSTKSTLGQTD 397

Qy 202 -----DGELEYLVKXKELSYDECYWESESDI-STFQNEIQRFKDVNSR----TRR 246

Db 398 FPAHSRKPAAPSNEPEYLCWMGLPYSECSWEDEALIGKFKQNCIDSFHSRNNSKTIPTRE 457

Qy 247 SKVDHKNRPRDFQQFDHTPEELKG--LLHPYQLEGLNPLRFSWSKQTHVILADEMGLG 303

Db 458 CKAL--KQRP-R-FVALKKQPAYLGGENLELDYQLEGLNLAHSAWCKNNSVILADEMGLG 514

Qy 304 KTIQSIALLASLFEENLI--PHLVIAPLSTLRNWEREFATWAPQMNVVMYFGTAQARAVI 361

Db 515 KTIQTSFLSYLFHQHQLYGPFLIVVPLSTLTWSQREFEWAPEINVVVYIGDLMsrNTI 574

Qy 362 REHEFYLSKDQKIKKKKSGQISSESQKRIKFDVLTYSYEMINLDSAVLKPIKWECEMIV 421

Db 575 REYEWI-----HSQTKRLKFNALITTYEILLKDKTVLGSINWAF LGV 616

Qy 422 DECHLKNKDSKLFSSLTQYSSNHRILLTGTPLONNLDELFMHFLDAGKFGSLEEFQE 481

Db 617 DEAHLKNDSDSLLYKTLIDFKSNHRLITGTPLQNSLKELWSLJHFIMPEKFEFWEDEE 676

Qy 482 EFKDINQEEQISRLHKMLAPHLRRVKKDVMKMPKKEILLRVDLSSLOKEYYKAIFTR 541

Db 677 DHGK-GRENGYQSLHKVLEPFLRRVKKDVEKSLPAKVEQILRVEMSAQKQYKWLITR 735

Qy 542 NYQVLTG-KGGAQISLNNIMMELRKVCCHPYMLEGVEP-VIHDANEAFKQLLESCGKLQL 599

Db 736 NYKALAKGTRGSTGFLNIVMELKCCNHCYLIKPEENERENGQEILLSLIRSGKLIL 795

Qy 600 LDKMMVKLKEQGHRLIYTFQFHMLDLLEDYCTHKKWQYERIDGKVGGAERQIRIDRFNA 659

Db 796 LDKLLTRLRERGNRVLIFSQVMRMLDILAEYLTIKHYPFQRLDGSIKGEIRKQALDHENA 855

Qy 660 KNSNKFCLLSTRAGGLGINLATADTVIYDSDWNPHADLOAMARAHRLGQTNKVMYRL 719

Db 856 DGSEDFCLLSTRAGGLGINLASADTVVIFDSWNPQNDLQAQARAHRIQKQKVNIYRL 915

Qy 720 INRGTIERRMMQLTKKKMVLHLVV-----GKLTQINQEEELDDIIRYG 764

Db 916 VTKGTVEEIEIERAKKQWLDHLVIQRMDDTTGRTILENNSRSNPNFENKEELTAILKFG 975

Qy 765 SKELFASDDDEAGSGKIHYYDAAIDKLLDRDLVEAEVSVDDDEENGFLKAFKVANFEY 824

Db 976 AEDLFKELEGESEP-----QEMDIDEIL--RLAETRENEVSTSATDELLSQFKVANFAT 1028

Qy 825 IDENEAALAEQARVAAESKSSAGNSDRASYWEELL-----KDKFELHQAEELNALGK 876

Db 1029 MEDEE-ELEER-----PHKD----WDEIIPEEKVKVEEERQKELEEIYMLPR 1072

Qy 877 RKRSRKQLVSI EEDDLAGLEDVSSDGDESYEAE-STDGEAAGQGVQTRRRPYRRKGRDNL 935

Db 1073 IRSSTKK-----AQTNDSDSDTESKRQAQRSSASESETEDDDDKKPKRR----- 1117

Qy 936 EPTPLMEGEGRSFR--VLGFNQSORAIFVQTLRMRYGAGNFDWKEFVPR--LKQKTFEE 989

Db 1118 -----GRPRSVRKDLVEGFTDAEIRRFIKAYKKFGL-PLERLECIARDAELVDKSVAD 1169

Qy 990 INEYGILF-----LKHIAEEIDENSPTFSFGVPKEGLRIEDVLVRIALLILVQEKVKF 1042

Db 1170 LKRLGELIHNCSVSAMQYEEQLKENASEGKPGKRRGPTTIKISGVQNVVKSIIQHEEFP 1229

Qy 1043 VEDHPGKPVFPFSR-----ILERFPGLRSGKIWKEEHDKIMTRAVLKHGYGRWQAIIVDDKE 1097

Db 1230 EMLHKSIPVDPEEKKYCLTCRVKAAHFVVEGWVEDDSRLLLGIYEHGYGNWELIKTDPE 1289

Qy 1098 LGIQELIIC-----KELNFPHISLSAAEQAGLQNGSGGSNPGAQTNQNPGSVIT 1147

Db 1290 LKLTDKILPVETDKKPGKQLQTRADYLLKLLRKGLEKK-----GAVT 1332

Qy 1148 GNNNASADGAQVNSMFYYRDMQRRLEVFVKRVLLEKAMNVEYAEYYGLGSSSIPT 1207

Db 1333 GGEEK-----LKKRKPVRKKNKVPRLKEEHGIELSSPRHSD 1370

Qy 1208 EPEAEPKIADTVGSFIEVDDEMLDGLPKTDPITSEEIMGAADVNNQARVEIAQHYNQMC 1267

Db 1371 NPSEEGEVKD-----DGLEKS-PMKKQKKKENKENKE-----KQMS 1406

Qy 1268 KLLD-ENAPESVQAYVNNQPPSTKVNESFRALKSINGNINTILSITSDQSKSHEDDTKPD 1326

Db 1407 SRKDKGDKERKKSDDKKEPKSGDAKSSSKSRSQGPVHITAGSEPVPIGEDEDD---D 1463

Qy 1327 LNNVMKDTAETKPLRGVVDLNVVEGEENIAE 1360

Db 1464 LDQETFSICKERMRPVKKALKQLDKPKDGLNVQE 1497

RESULT 5

US-09-535-008-67

Query Match 14.3%; Score 1028.5; DB 4; Length 1646;

Best Local Similarity 30.8%; Pred. No. 1e-76;

Matches 300; Conservative 156; Mismatches 324; Indels 195; Gaps 32;

Qy 269 LKGLLHPYQLEGLNFLRFSWSKQTHVILADEMGLGKTIQSIALLASLFEENLI--PHLVI 326

Db 750 VNGVLKQYQIKGLEWLVSLYNNNLGILADEMGLGKTIQTIATILYLMHKRINGPFLII 809

Qy 327 APLSTLRNWEREFATWAPQMNVMYFGTAQARAVIREHEFYLSKDKKIKKKSQGISSE 386

Db 810 VPLSTLSNWAYEFDKWAPSVMKSYKSPAARRAF-----VPQLRSG----- 851

Qy 387 SKQRIKIFDVLITSYEMINLDSAVLKPKECMIVDEGHLRKNKDSKLFSSL-TQYSSNH 445

Db 852 -----KENVLLTTYEYIIKDKHILAKIRWKYMIVDEGHRMKNHHCKLTQVLNTHYVAPR 905

Qy 446 RILLTGTPLONNLDELFMLMHFMDAGKFGSLEEFQEEFK-----DINQEEQ---I 492

Db 906 RLLLTGTPLONKLPQLWALLNLLPTIFKSCSTFEQWFNAPPANTGEKVVDLNEETILII 965

Qy 493 SRLHKMLAPHLRRVKQDMKMPKKELILRVDLSSLQKEYY-----KAIFTRNYQVLT 547

Db 966 RRLHKVLRPFLRLRLLKKEVEAQLPEKVEYVIKCDMSALQRLVYRHMQAQGVLLTDGSEKD 1025

Qy 548 KKG-GAQISLNNIMMELRKVCCHPYMLEGVEPVIHDANEAFKQ-----LLES 593

Db 1026 KKGKGGTKTLMNTIMQLRKICNHPYMFQIE-----ESFSEHLGFTGIVQGLDLYRA 1078

Qy 594 CGKQLQLDKMVKLKEQGRVLIYTOFQHMLDLLEDYCTHKWQYERIDGKVGGAERQIR 653

Db 1079 SGKFELLDRILPKLRATNHNKVLFLFCQMTSLMTIMEDYFAYRGFKYLRDLDTGTTKAEDRGML 1138

Qy 654 IDRFNAKNSNKFCLLSTRAGGLGINLATADTVIIYSDWNPHADLOAMARHRLGQTNK 713

Db 1139 LKTFNEPGSEYFIFLLSTRAGGLGLNLQSAADTVIIFSDWNPHQDLOAQDRAHRIGQQNE 1198

Qy 714 VMIYRLINRGTIERMMQLTKKKQVLEHLVV--GKLKTQINQEE---LDDIIRYGSKEL 768

Db 1199 VRVLRLLCTVNSVEEKILAAAKYKLNVDQKVIQAGMFDQKSSSHERRAFLQAILEH----- 1253

Qy 769 FASEDDE-----AGKSGKIHY--DDAAIDKLLDRDLVEAEVSVVDEEENGFLKAFKV 819

Db 1254 --EEQDESRLHCSTGSGSASFAHTAPPPAGVNPDLPEEPPLKBEDEVDPDETVMQMIARHE- 1310

Qy 820 ANFEYIDENEAALAEQ RVAES--KSSAGNSDRASYWELLKDKFELH---QAEELN 872

Db 1311 ---EEFDLFMRMDLDRRREEARNPKRKPRLMEEDELPSW--IIKDDAEVERLTCEEEEEK 1365

Qy 873 ALGKRKRSRKQ-----LVSIEEDDLAGLEDV-----SSDGEDSYEA 908

Db 1366 MFGGSRHRKEVDYSDSLTEKQWLKAIEEGTLEEIEEEVQKKSSRKRKRDS DAGSSTPT 1425

Qy 909 EST-----DGEAAGQGVQTRRRPYRRKGR---DNLEPT-----LMEGE 944

Db 1426 TSTRSRDKDDESKQ-----KKGRPPAEKLSNPNNLTKKMKKIYDAVIKYDSS 1476

Qy 945 GRSFRVLGFNQSORAIFVQTLNRYGAGNF-----DWKEFVPRLKQKTFEEINEYGIL 996

Db 1477 GRQL-----SEVFIQLPSRKELPEYYELIRKPVDFKKIKERIRNHNKYSRLND---- 1523

Qy 997 FLKHIAEEIDENSPTFSDGVPKEG-LRIEDVLVRIALLILVQEKVKFVEDHPGKVPFPSR 1055

Db 1524 -LEKDVMLLCQNAQTFN---LEGSLIYEDSIVLQSVFTSVRQKIEKEDDSEGE-----E 1573

Qy 1056 ILERFPGLRSGKIWKEEHDKIMIRAVLKHGYGRWQAIYDDKELGIQELICKELNFPHISL 1115

Db 1574 SEEEEEGEEGSESESRSVKVIK-----LGRKEKAQRLKGRRRPSRGRAPVVSVD 1627

Qy 1116 SAAEQAGLQGGNGSG 1130

Db 1628 DDSEEEQEEEDRSGSG 1642

RESULT 6

US-09-535-008-2

Query Match 14.3%; Score 1027; DB 4; Length 1647;

Best Local Similarity 30.9%; Pred. No. 1.3e-76;

Matches 301; Conservative 157; Mismatches 325; Indels 190; Gaps 32;

Qy 269 LKGLLHPYQLEGLNFLRFSWSKQTHVILADEMGLGKTIQSIALLASLFEENLI--PHLVI 326

Db 750 VNGVLKQYQIKGLEWLVSLYNNNLGILADEMGLGKTIQTIATILYLMHKRINGPFLII 809

Qy 327 APLSTLRNWEREFATWAPQMNVMYFGTAQARAVIREHEFYLSKDKKIKKKSQGISSE 386

Db 810 VPLSTLSNWAYEFDKWAPSVMKSYKSPAARRAF-----VPQLRSG----- 851

Qy 387 SKQRIKIFDVLITSYEMINLDSAVLKPKECMIVDEGHLRKNKDSKLFSSL-TQYSSNH 445

Db 852 -----KENVLLTTYEYIIKDKHILAKIRWKYMIVDEGHRMKNHHCKLTQVLNTHYVAPR 905

Qy 446 RILLTGTPLONNLDELFMLMHFMDAGKFGSLEEFQEEFK-----DINQEEQ---I 492

Db 906 RLLLTGTPLONKLPQLWALLNLLPTIFKSCSTFEQWFNAPPANTGEKVVDLNEETILII 965

Qy 493 SRLHKMLAPHLRRVKQDMKMPKKELILRVDLSSLQKEYY-----KAIFTRNYQVLT 547

Db 966 RRLHKVLRPFLRLRLLKKEVEAQLPEKVEYVIKCDMSALQRLVYRHMQAQGVLLTDGSEKD 1025

Qy 548 KKG-GAQISLNNIMMELRKVCCHPYMLEGVEPVIHDANEAFKQ-----LLES 593

Db 1026 KKGKGGTKTLMNTIMQLRKICNHPYMFQIE-----ESFSEHLGFTGIVQGLDLYRA 1078

Qy 594 CGKQLQLDKMVKLKEQGRVLIYTOFQHMLDLLEDYCTHKWQYERIDGKVGGAERQIR 653

Db 1079 SGKFELLDRILPKLRATNHNKVLFLFCQMTSLMTIMEDYFAYRGFKYLRDLDTGTTKAEDRGML 1138

US-09-535-008-2

Sequence 2, Application US/09535008

Patent No. 6465629

GENERAL INFORMATION:

APPLICANT: Wong, Alexander K.C.

APPLICANT: Tavtigian, Sean V.

APPLICANT: Teng, David H.-F.

TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE

TITLE OF INVENTION: AND OTHER CANCER TYPES

FILE REFERENCE: 2318-259

CURRENT APPLICATION NUMBER: US/09/535,008

CURRENT FILING DATE: 2000-03-23

EARLIER APPLICATION NUMBER: U.S. 60/125,806

EARLIER FILING DATE: 1999-03-23

NUMBER OF SEQ ID NOS: 77

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 2

LENGTH: 1647

TYPE: PRT

ORGANISM: Homo sapiens

US-09-535-008-2

Query Match 14.3%; Score 1027; DB 4; Length 1647;

Best Local Similarity 30.9%; Pred. No. 1.3e-76;

Matches 301; Conservative 157; Mismatches 325; Indels 190; Gaps 32;

Qy 269 LKGLLHPYQLEGLNFLRFSWSKQTHVILADEMGLGKTIQSIALLASLFEENLI--PHLVI 326

Db 750 VNGVLKQYQIKGLEWLVSLYNNNLGILADEMGLGKTIQTIATILYLMHKRINGPFLII 809



QY	327	APLSTLRNWEREFATWAPQMNVMYFGTAQARAVIREHEFYLSKQDKIKKKKSGQISSE	386
Db	810	VPLSTLSNWAYEFDKWAPSVVKVSYKGSPPAARAF-----VPQLRSG-----	851
QY	387	SKQKRIKFDVLTYSEMINLDSAVLKPIKWECMIVDEGHRLLKNKOSKLFSSIL-TQYSSNH	445
Db	852	-----KFNVLTTYEYIIKDKHILAKIRWKYMIIVDEGHRMKNHHCKLTQVLNTHYVAPR	905
QY	446	RILLTGTPLONNLDELFWLMHFLDAGKFGSLEEFQEEFK-----DINQEEQ---I	492
Db	906	RLLLTGTPLONKLPFLWALLNFLPTIFKSCSTFEQWFNAPFAMTGEKVLDNEEETILII	965
QY	493	SRLHKMLAPHLRRVKDVMDMPKPKELILRVDLSSLOKEYY-----KAIFTRNYQVLT	547
Db	966	RRLHKVLRPFLRLRLKKEVEAQLPEKVEYVIKCDMSALQORVLYRHMQAQGVLLTDGSEKD	1025
QY	548	KKG-GAQISLNNIMELRKVCCHPYMLEGVEPVIHDANEAFKQ-----LLES	593
Db	1026	KKGKGGTKTLMNTINQLRKICNHPYMFQHIE-----ESPSEHLGFTGGIVQGLDLYRA	1078
QY	594	CGKLQLLDKMMVKLKEQGHRVLIYTQFQHMLDLLEDYCTHKKWQYERIDKGVGGAERQIR	653
Db	1079	SGKFELDLRILPKLRATNHKVLFCQMTSLMTIMEDYFAYRGFKYLRLDGTTKAEDRGML	1138
QY	654	IDRFNAKNSNKFCLLSTRAGGLGINLATADTVIIYSDWNPHADLQAMARHRLGQTNK	713
Db	1139	LKTFNEPGSEYFIFLLSTRAGGLGLNLQSAADTVIIFSDWNPHQDLQAQDRAHRIGQONE	1198
QY	714	VMIYRLINRGTIERMMQLTKKQWVLEHLV--GKLKTQINQEE--LDDIIRYGSKEL	768
Db	1199	VRVLRLLCTVNSVEEKILAAAKYKLVNDQKVIQAGMFDQKSSSHERRAFLQAILEH-----	1253
QY	769	FASEDDE-----ACKSGKIHY--DDAAIDKLDRDLVEAEVSDDEENGFLKAFKV	819
Db	1254	--EEQDESRLHCSTGSGSASFAHTAPPAGVNPDLLEPPKKEEDEVDPDETVMQIARHE-	1310
QY	820	ANFEYIDENEAALAEARVAES--KSSAGNSDRASIWEELLKKKFLH----QAEBLN	872
Db	1311	---EEFDLFMRMDLDRRREEARNPKRKPRLMEEDLPSW--IIKDDAEVERLTCEEEEEK	1365
QY	873	ALGKRKRSRKQ-----LVSIEEDDLAGLEDV-----SSDGDESIEYA	908
Db	1366	MFGGRGSRHRKEVDYDSLTEKQWLKAIEEGTLEETIEEEVROKSSRRKRKRDSDAGSSTPT	1425
QY	909	EST-----DGEAAGQGVQTRRRPYRRKGR---DNLEPTP-----LMEGEGR	946
Db	1426	TSTRSRDKDDESKQ-----KGRGRPPAEKLSNPPNLTKKMKKIVDAVIKYKDSS	1476
QY	947	SFRVLGFNQSORAIFVOTLMRYGAGNF-----DWKEFVPRLKQKTFEEINEYGILFL	998
Db	1477	SGRQLS-----EVFIQLPSRKELPEYYELIRKPVDFFKKIKERIRNHKYRSLND-----L	1525
QY	999	KHIAEEIDENSPTFSDGVPKEG-LRIEDVLVRIALLILVQEKVFEVDHPGKVPFPSPRIL	1057
Db	1526	EKDVMLLCQNAQTFN----LEGSLIYEDSVLQSVFTSVRQKIEKEDDSEGE-----ESE	1576
QY	1058	ERFPGLRSKGIWKEEHDKIMIRAVLKHGYGRWQAIVDDKELGIQELICKELNFPHISLSA	1117
Db	1577	EEEEEGEEGSESESRSVKVIK-----LGRKEKAQDRLLKGGRRRPSRGRAPVVSDDD	1630
QY	1118	AEQAGLQONGSG	1130
Db	1631	SEEEOEEDRS	1643

## RESULT 7

US-09-535-008-75

03-03-333-008-73  
; Sequence 75, Application US/09535008

; Patent No. 6465629

GENERAL INFORMATION:

APPLICANT: Wong, Alexander K.C.

APPLICANT: Tavtigian, Sean V.

APPLICANT: Teng, David H.-F.

Db 1477 DSSGRQL-----SEVFIQLPSRKELPEYYELIRKVPDFKKIKERIRNHKYSRLND- 1526

Qy 994 GILFLKHIAEIDENSPTFSDGVPKEG-LRIEDVLVRIALLILVQEKVKFVEDHPGKPVF 1052

Db 1527 ----LEKDWMLLCQNAQTFN----LEGLSIYEDSIVLQSVFTSVRQKIEKEDDSEGE--- 1575

Qy 1053 PSRILFRPGLRSGKIWKEEHDKIMIRAVLKHGGRWQAIVDDKELGIQELICKELNFPH 1112

Db 1576 --ESEEEEEEGEESERSVKVKK-----LGRKEKAQDLKGGRRRPSRGSRAKPV 1627

Qy 1113 ISLSAAEQAGLQONGSG 1130

Db 1628 VSDDSEEEQEEDRSGSG 1645

RESULT 8

US-09-535-008-71

; Sequence 71, Application US/095335008

; Patent No. 6465629

; GENERAL INFORMATION:

; APPLICANT: Wong, Alexander K.C.

; APPLICANT: Tavtigian, Sean V.

; APPLICANT: Teng, David H.-F.

; TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE

; FILE REFERENCE: 2318-259

; CURRENT APPLICATION NUMBER: US/09/535,008

; CURRENT FILING DATE: 2000-03-23

; EARLIER APPLICATION NUMBER: U.S. 60/125,806

; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 71

; LENGTH: 1650

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-535-008-71

Query Match 14.3%; Score 1025.5; DB 4; Length 1650;

Best Local Similarity 30.8%; Pred. No. 1.8e-76;

Matches 301; Conservative 157; Mismatches 325; Indels 193; Gaps 32;

Qy 269 LKGLLHPYQLEGLNFLRFSWSKQTHVILADEMGLGKTIQSIALLASLFEENLI--PHLVI 326

Db 750 VNGVLKQYQIKGLEWLVSLYNNNLGILADEMGLGKTIQTIALTITYLMEHKRINGPFLII 809

Qy 327 APLSTLRNWEREFATWAPQMNVMYFGTAQARAVIREHEFEYLSKDQKKIKKKSGQISSE 386

Db 810 VPLSTLSNWAYEFDKWAPSVVKVYKGSPPAARRAF-----VPQLRSG----- 851

Qy 387 SKQRIKFDVLLTSYEMINLDSAVLKPIKWECEMIVDEGHRKLNKDSKLFSSL-TQYSSNH 445

Db 852 -----KFNVLTTYEYIIKDHILAKIRWKYMIWDEGHRMKNHCKLTQVLNTHYVAPR 905

Qy 446 RIILTTGPLQNNLDELFMLMHFLDAGKFGSLEEFQEEFK-----DINQEEQ---I 492

Db 906 RLILLTGTPQLNKLPELWALLNLLPFTIFKSCSTFEQWFNAPFAMTGEKVDLNEETILII 965

Qy 493 SRLHKMLAPHLLRRVKDKVMDMPKKELILRVDLSSLQKEY-----KAIFTRNYQVLT 547

Db 966 RRLHKVLRPFLRLRLKKEVEAQLPEKVEYVIKCDMSALQRLVLRHMQAKGVLLTDGSEKD 1025

Qy 548 KKG-GAQISLNNIMMELRKVCCHPYMLEGVEPVIHDANEAFKQ-----LLES 593

Db 1026 KKGKGGTKTLMNTIMQLRKICNHPYMFQHIE-----ESPSEHLGFTGGIVQGLDLYRA 1078

Qy 594 CGKLQLLDKMMVKLKEQGRVLIYTFQHMLDLLEDYCTHKKQYERIDGKVGGAERQIR 653

Db 1079 SGKFELLDRIPLKLRATNHNKVLFCQMTSLMTIMEDYFAYRGPVKYLRLDGTTKAEDRGML 1138

Qy 654 IDRFNAKSNKFCFLLSTRAGGLGINLATADTVIIYDSDNPHADLQAMARHLGQTNK 713

Db 1139 LKTFNEPGEYFIFLLSTRAGGLGILNQSADTVIIFSDWNPHQDLQAQDRAHRIGQONE 1198

Qy 714 VMIYRLINRGTIERMMQLTKKMMVLEHLV--GKLKTQINQEE---LDDIIRYGSKEL 768

Db 1199 VRVLRLLCTVNSVEEKILAAAKYKLNVDQKVIQAGMFDQSSSHERRAFLQAILHEH----- 1253

Qy 769 FASEDDE-----AGSGKIHY--DDAAIDKLLDRDLVEABEVSVDDDEENGFLKAFKV 819

Db 1254 --EEQDESRCSTGSGSASFAHTAPPAGVNPDLPEPPLKEBEVDPDETVMQMIARHE- 1310

Qy 820 ANFEYIDENEAALAEARVAAES---KSSAGNSDRASYWEELLKDKFELH---QAEELN 872

Db 1311 --EEFDLFMRMDLDRRREEARNPKRKPRLMEEDELPSW--IIKDDAEVERLTCEEEEEK 1365

Qy 873 ALGKRKRSRKQ-----LVSIEEDDLAGLEDV-----SSDGDES 905

Db 1366 MFGRGSRHRKEVDYSDSLTEKQWLKTLKAIEEGTLEIEIEEVRQKSSRKRRDSDAGSS 1425

Qy 906 YEAEST-----DGEAAGQGVQTGRRPYRRKGR--DNLEPTP-----LMEG 943

Db 1426 TPTTSTRSRDKDDESKQ-----KXGRPPPAEKLSPNPNLTKKMKKIYDAVIKYK 1476

Qy 944 EGRSFRVLGFNQSORAIFVQTLNRYGAGNF-----DWKEFVPRLLKQKTFEEINEYGI 995

Db 1477 DSSSGRQLS-----EVFIQLPSRKELPEYYELIRKVPDFKKIKERIRNHKYSRLND--- 1527

Qy 996 LFLKHIAEIIDENSPTFSDGVPKEG-LRIEDVLVRIALLILVQEKVKFVEDHPGKPVFPS 1054

Db 1528 --LEKDVMLLCQNAQTFN----LEGLSIYEDSIVLQSVFTSVRQKIEKEDDSEGE----- 1576

Qy 1055 RILFRFPGLRSGKIWKEEHDKIMIRAVLKHGGRWQAIVDDKELGIQELICKELNFPHIS 1114

Db 1577 ESEEEEEEGEESERSVKVKK-----LGRKEKAQDLKGGRRRPSRGSRAKPVVS 1630

Qy 1115 LSAAEQAGLQONGSG 1130

Db 1631 DDDSEEEQEEDRSGSG 1646

RESULT 9

US-09-535-008-77

; Sequence 77, Application US/095335008

; Patent No. 6465629

; GENERAL INFORMATION:

; APPLICANT: Wong, Alexander K.C.

; APPLICANT: Tavtigian, Sean V.

; APPLICANT: Teng, David H.-F.

; TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE

; FILE REFERENCE: 2318-259

; CURRENT APPLICATION NUMBER: US/09/535,008

; CURRENT FILING DATE: 2000-03-23

; EARLIER APPLICATION NUMBER: U.S. 60/125,806

; EARLIER FILING DATE: 1999-03-23

; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 77

; LENGTH: 1681

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-535-008-77

Query Match 14.1%; Score 1011; DB 4; Length 1681;

Best Local Similarity 29.7%; Pred. No. 3.1e-75;

Matches 300; Conservative 156; Mismatches 324; Indels 230; Gaps 32;

Qy 269 LKGLLHPYQLEGLNFLRFSWSKQTHVILADEMGLGKTIQSIALLASLFEENLI--PHLVI 326

Db 750 VNGVLKQYQIKGLEWLVSLYNNNLGILADEMGLGKTIQTIALTITYLMEHKRINGPFLII 809

Qy 327 APLSTLRNWEREFATWAPQMNVMYFGTAQARAVIREHEFEYLSKDQKKIKKKSGQISSE 386

Db 810 VPLSTLSNWAYEFDKWAPSVVKVYKGSPPAARRAF-----VPQLRSG----- 851

QY 387 SKQRIKFDVLLTSYEMINLDSAVLKPWKECMIVDEGHRLLKNKDSKLFSSL-TOYSSNH 445  
Db 852 -----KFNVLTTYEYIIKDKHILAKIRWKYMIVDEGHRMKNHCKLTQVLNTHYVAPR 905  
QY 446 RILLTGTPLONNLDELFMHFLDAGKFGSLEEFQEFK-----DINQEEQ---I 492  
Db 906 RLLLTGTPLQNKLPFWALLNLLPTIFKSCSTFEQWFNAPFAMTGEKVDLNEEETILII 965  
QY 493 SRLHKMLAPHLRRVKDKVMKMPKPKELILRVDSLLOKEYY-----KAIFTRNYQVLT 547  
Db 966 RRLHKVLRPFLLRRLKKEVEAQLPEKVEYVIKCDMSALQRLVYRHMQAQGVLLTDGSEKD 1025  
QY 548 KKG-GAQISLNNIMMELRKVCCHPYMLEGVEPVIHDANEAFKQ-----LLES 593  
Db 1026 KKGKGGTKLMTNTIMQLRKICNHPYMFQHIE-----ESFSEHLGFTGGIVQGLDLYRA 1078  
QY 594 CGKLQLLDKMMVKLKEQHRVLIYTQFQHMDDLLEDYCTHKKWQYBRIDGKVGGAERQIR 653  
Db 1079 SGKFELLDRIPLKLRATNHKVLFCQMTSLMTIMEDYFAYRGFKYLRLDGTTKAEDRGML 1138  
QY 654 IDRFNAKNSKFCFLLSTRAGGLGINLATADTVIIYDSDNPHADLQAMARAHRLGQTNK 713  
Db 1139 LKTNEPGSEYFIFLLSTRAGGLGNLQSDATVIIFSDWNPHQDLQAQDRAHRIGQQNE 1198  
QY 714 VMIYRLINRGTTIEERMMQLTKKMWLEHLVV--GKLKTQINQEE---LDDIIRYGSSEL 768  
Db 1199 VRVLRCTVNSVEEKILAAKYKLVNDQKVIQAGMFDQKSSSHERRAFLQAILEH----- 1253  
QY 769 FASEDDE-----AGSGKIHY--DDAAIDKLLDRDLVEAEVSVDEEENGFLKAFKV 819  
Db 1254 --EEQDESRLCTVNSVEEKILAAKYKLVNDQKVIQAGMFDQKSSSHERRAFLQAILEH----- 1310  
QY 820 ANFEYIDENEAAALEAQRVAES---KSSAGNSDRASYWEELLLKDKFELH---QAEELN 872  
Db 1311 ---EEFDLFMRMDLDRRREEARNPVKPRLMEEDLPSW--IIKDDAEVERLTCEEEEEK 1365  
QY 873 ALGKRKRSRKQ-----LV 885  
Db 1366 MFGRGSRHRKEVDYDSLTEKQWLKKITGKDIDHTASSVARGLQFORGLQFCTRASKTLK 1425  
QY 886 SIEEDDLAGLEDV-----SSDGDSEYEAEST-----DGEAAGQGVQGTGRRPYR 928  
Db 1426 AIEEGTLEEIEEVRQKSSRKRRKRDSDAGSSTPTTSTRSRDKDDESKKQ-----K 1476  
QY 929 RKGR---DNLEPT-----LMEGEGRSFRVLGFGNQSQRALFVQTLMRYG 969  
Db 1477 KRGRPPAEKLSPPNPNTKKMKKIYDAVIKYDSSGRQL-----SEVFIQLPSRKE 1527  
QY 970 AGNF-----DWKEFVPRLKQKTFFEEINEYGILFLKHIAEEIDENSPTFSDGVPKEG- 1020  
Db 1528 LPEYYELIRKVPDFKIKERIRNHKYRSLND-----LEKDVMLLCQNAQTFN----LEGS 1578  
QY 1021 LRIEDVLVRIALLIVQEKVKFVEDHPGKVPFPPSRILERPGLRSGKIWKBEHDKIMIRA 1080  
Db 1579 LIYEDSIVLQSVFTSVRQKIEKEDDSEGE-----ESEEEEEEGEESRSRQKVKIK- 1632  
QY 1081 VLKHGYGRWQAIVDKELGIQELICKELNFPHISLSAAEQAGLQONGSG 1130  
Db 1633 -----LGRKEKAQDRLKGGRRRPSRGSRAKPVVSDDDSEEEQEDRSKSG 1677

RESULT 10  
US-09-535-008-69  
; Sequence 69, Application US/09535008  
; Patent No. 6465629  
; GENERAL INFORMATION:  
; APPLICANT: Wong, Alexander K.C.  
; APPLICANT: Tavtigian, Sean V.  
; APPLICANT: Teng, David H.-F.  
; TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE  
; TITLE OF INVENTION: AND OTHER CANCER TYPES  
; FILE REFERENCE: 2318-259  
; CURRENT APPLICATION NUMBER: US/09/535,008

; CURRENT FILING DATE: 2000-03-23  
; EARLIER APPLICATION NUMBER: U.S. 60/125,806  
; EARLIER FILING DATE: 1999-03-23  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 69  
; LENGTH: 1678  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-535-008-69  
  
Query Match 14.0%; Score 1009.5; DB 4; Length 1678;  
Best Local Similarity 29.7%; Pred. No. 4.1e-75;  
Matches 299; Conservative 157; Mismatches 324; Indels 227; Gaps 32;  
  
QY 269 LKGLLHPYQLEGLNFLRFSWSKQTHVILADEMGLGKTIQSIALLASLFEENLI--PHLVI 326  
Db 750 VNGVLKQYQIKGLEWLVSLYNNLNLGILADEMGLGKTIQTIATILYLMHKRINGPFLII 809  
QY 327 APLSTLRNWEREFATWAPQNMVVMYFGTAQARAVIREHEFYLSKDQKIKKKKSGQISSE 386  
Db 810 VPLSTLSNWAYEFDKWAPSVVKVSGSPAARAF-----VPQLRSG----- 851  
QY 387 SKQRIKFDVLLTSYEMINLDSAVLKPWKECMIVDEGHRLLKNKDSKLFSSL-TOYSSNH 445  
Db 852 -----KFNVLTTYEYIIKDKHILAKIRWKYMIVDEGHRMKNHCKLTQVLNTHYVAPR 905  
QY 446 RILLTGTPLONNLDELFMHFLDAGKFGSLEEFQEFK-----DINQEEQ---I 492  
Db 906 RLLLTGTPLQNKLPFWALLNLLPTIFKSCSTFEQWFNAPFAMTGEKVDLNEEETILII 965  
QY 493 SRLHKMLAPHLRRVKDKVMKMPKPKELILRVDSLLOKEYY-----KAIFTRNYQVLT 547  
Db 966 RRLHKVLRPFLLRRLKKEVEAQLPEKVEYVIKCDMSALQRLVYRHMQAQGVLLTDGSEKD 1025  
QY 548 KKG-GAQISLNNIMMELRKVCCHPYMLEGVEPVIHDANEAFKQ-----LLES 593  
Db 1026 KKGKGGTKLMTNTIMQLRKICNHPYMFQHIE-----ESFSEHLGFTGGIVQGLDLYRA 1078  
QY 594 CGKLQLLDKMMVKLKEQHRVLIYTQFQHMDDLLEDYCTHKKWQYBRIDGKVGGAERQIR 653  
Db 1079 SGKFELLDRIPLKLRATNHKVLFCQMTSLMTIMEDYFAYRGFKYLRLDGTTKAEDRGML 1138  
QY 654 IDRFNAKNSKFCFLLSTRAGGLGINLATADTVIIYDSDNPHADLQAMARAHRLGQTNK 713  
Db 1139 LKTNEPGSEYFIFLLSTRAGGLGNLQSDATVIIFSDWNPHQDLQAQDRAHRIGQQNE 1198  
QY 714 VMIYRLINRGTTIEERMMQLTKKMWLEHLVV--GKLKTQINQEE---LDDIIRYGSSEL 768  
Db 1199 VRVLRCTVNSVEEKILAAKYKLVNDQKVIQAGMFDQKSSSHERRAFLQAILEH----- 1253  
QY 769 FASEDDE-----AGSGKIHY--DDAAIDKLLDRDLVEAEVSVDEEENGFLKAFKV 819  
Db 1254 --EEQDESRLCTVNSVEEKILAAKYKLVNDQKVIQAGMFDQKSSSHERRAFLQAILEH----- 1310  
QY 820 ANFEYIDENEAAALEAQRVAES---KSSAGNSDRASYWEELLLKDKFELH---QAEELN 872  
Db 1311 ---EEFDLFMRMDLDRRREEARNPVKPRLMEEDLPSW--IIKDDAEVERLTCEEEEEK 1365  
QY 873 ALGKRKRSRKQ-----VSIE 888  
Db 1366 MFGRGSRHRKEVDYDSLTEKQWLKKITGKDIDHTASSVARGLQFORGLQFCTRASKAIE 1425  
QY 889 EDDLGLAGLEDV-----SSDGDSEYEAEST-----DGEAAGQGVQGTGRRPYRRKG 931  
Db 1426 EGTLEEIEEVRQKSSRKRRKRDSDAGSSTPTTSTRSRDKDDESKKQ-----KGRG 1476  
QY 932 R---DNLEPT-----LMEGEGRSFRVLGFGNQSQRALFVQTLMRYGAGN 972  
Db 1477 RPPAEKLSPPNPNTKKMKKIYDAVIKYDSSGRQL-----SEVFIQLPSRKELPE 1527  
QY 973 F-----DWKEFVPRLKQKTFFEEINEYGILFLKHIAEEIDENSPTFSDGVPKEG-LRI 1023



Db 1528 YYELIRKPVDFKKIKERIRNHKYSRLND-----LEKDVMLLCQNAQTFN-----LEGSLIY 1578

Qy 1024 EDVLVRIALLILVQEKVKFVEDHPGKPVFPSPRILERFPGLSRSGKIWKEEHDKIMIRAVLK 1083

Db 1579 EDSIVLQSVFTSVRQKIEKEDDSEGE-----ESEEEEEEGEESSESRSVKVLIK-----1629

Qy 1084 HGYGRWOAIVDDKELGIGIQLICKELNFPFPHISLSAAEQAGLQONGSG 1130

Db 1630 --LGRKEKAQDRLKGGRRRPSRGSRAKPVVSDDDSEEEQEEEDRSGS 1674

RESULT 11

US-09-535-008-73

; Sequence 73, Application US/095335008

; Patent No. 6465629

; GENERAL INFORMATION:

; APPLICANT: Wong, Alexander K.C.

; APPLICANT: Tavtigian, Sean V.

; APPLICANT: Teng, David H.-F.

; TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE

; TITLE OF INVENTION: AND OTHER CANCER TYPES

; FILE REFERENCE: 2318-259

; CURRENT APPLICATION NUMBER: US/09/535,008

; CURRENT FILING DATE: 2000-03-23

; EARLIER APPLICATION NUMBER: U.S. 60/125,806

; EARLIER FILING DATE: 1999-03-23

; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 73

; LENGTH: 1682

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-535-008-73

Query Match 14.0%; Score 1009.5; DB 4; Length 1682;

Best Local Similarity 29.9%; Pred. No. 4.1e-75;

Matches 301; Conservative 157; Mismatches 325; Indels 225; Gaps 32;

Qy 269 LKGLLHPYQLEGLNFRFSWSKQTHVILADEMGLGKTIQSIALLASLFEENLI--PHLVI 326

Db 750 VNGVLKQYQIKGLEWLVSLYNNLNGILADEMGLGKTIQTIOTALITYLMEHCRINGPFLII 809

Qy 327 APLSTLRNWEREFATWAPQMVVYFGTAQARAVIREHEFYLSKDQKIKKKSGQISSE 386

Db 810 VPLSTLSNWAYEFDKWAPSVVYKGSPPARRAF-----VPQLRSG-----851

Qy 387 SKQRIKFDVLLTSYEMINLDSAVLKPIKWECEMIVDEGHRLLKNKDSKLFSSL-TQYSSNH 445

Db 852 -----KPNVLLTTYEYIIKDKHILAKIRWKYMIVDEGHRMKNHCKLTQVLNTHYVAPR 905

Qy 446 RLLTGTPLQNNLDELFLMHFLDAGKFGSLEEFQEEFK-----DINQEEQ---I 492

Db 906 RLLTGTPLQNKLPQELWALLNLLPTIFPKSCSTFEQWFNAPFAMTGEKVDLNEETILII 965

Qy 493 SRLHKMLAPHLRLRVKQDVMDKMPKXELILRVDLSSLQKEY-----KAIFTRNYQVLT 547

Db 966 RRLHKVLRPFLRLRLKKEVEAQLPEKVEYVIKCDMSALQRLVYRHMQAQGVLLTDGSEKD 1025

Qy 548 KKG-GAQISLNNIMELRKVCCHPYMLEGVEPVIHDANEAFKQ-----LLES 593

Db 1026 KKGKGGYKTLNMTIMQLRKICNHPYMFQHIE-----ESFSEHLGFTGIVQGLDLYRA 1078

Qy 594 CGKLQLLDKMMVKLEQGHRLVIYTFQHMDDLLEDYCTHKKKQYERIDGKVGGAERQIR 653

Db 1079 SGKFEELDRILPKLRATNKHVLLFCQWTSMTIMEDYFAYRGEFKYLRDLGTTKAEDRGML 1138

Qy 654 IDRFNAKNSNKFCELLSTRAGGLGINLATADTVIIYSDWNPHADLQAMARHLGQTNK 713

Db 1139 LKTFNEPGSEYFIFLLSTRAGGLGNLQADTVIIFSDWNPHQDLQAQDRAHRIGQQNE 1198

Qy 714 VMIYRLINRGFTIERMMQLTKKQVLEHLVV--GKLKTQINQEE---LDDIIRYGSSEL 768

Db 1199 VRVLRCTVNSVEEKILAAAKYKLVNDQKVIQAGMFDQKSSSHERRAFLOALEH-----1253

Qy 769 FASEDD-----AGKSGKIHY--DDAAIDKLLDRDLVEAEEVSVDDDEENGFLKAFKV 819

Db 1254 --EEQDESRCSTGSGSASFAHTAPPPAGVNPDLPEPLKEDEVPDDETVNMQIARHE- 1310

Qy 820 ANFEYIDENEAAALEAQRVAES---KSSAGNSDRASVWEELLKDKFELH-----QAEELN 872

Db 1311 ---EEFDLFMRMDLDRRREEARNPKRKPRLMEEDELPSW--IIKDDAEVERLTCEEEEEK 1365

Qy 873 ALGKRKRSRKQ-----LV 885

Db 1366 MFGGSRHRKEVDYSDSLTEKQWLKKITGDIHDTASSVARGLQFQRLQFCTRASKTLK 1425

Qy 886 SIEEDDLAGLEDV-----SSDGDESYEAEST-----DGEAAGQGVQGTGRRPYR 928

Db 1426 AIEEGTLEEIEEEVRQKSSRKRKRSDAGSSTPTTSTRSRDKDDESKQ-----K 1476

Qy 929 RKGR---DNLEPTP-----LMEGEGRSFRVLGFNQSQRAIFVQTLRMRYGAG 971

Db 1477 XGRPPAEKLSPPNPLTKMKKIIVDAVIKYKDSSSGRQLS-----EVFIQLPSRKELP 1530

Qy 972 NF-----DWKEFVPRLKQKTFEEINEYGILFLKHTAEEIDENSPFSDGVPKEG-LR 1022

Db 1531 EYELIRKPVDFKKIKERIRNHKYSRLND-----LEKDVMLLCQNAQTFN-----LEGSLI 1581

Qy 1023 IEDVLVRIALLILVQEKVKFVEDHPGKPVFPSPRILERFPGLSRSGKIWKEEHDKIMIRAVL 1082

Db 1582 YEDSIVLQSVFTSVRQKIEKEDDSEGE-----ESEEEEEEGEESSESRSVKVLIK-----1633

Qy 1083 KHGYGRWOAIVDDKELGIGIQLICKELNFPFPHISLSAAEQAGLQONGSG 1130

Db 1634 ---LGRKEKAQDRLKGGRRRPSRGSRAKPVVSDDDSEEEQEEEDRSGS 1678

RESULT 12

US-09-535-008-65

; Sequence 65, Application US/095335008

; Patent No. 6465629

; GENERAL INFORMATION:

; APPLICANT: Wong, Alexander K.C.

; APPLICANT: Tavtigian, Sean V.

; APPLICANT: Teng, David H.-F.

; TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE

; TITLE OF INVENTION: AND OTHER CANCER TYPES

; FILE REFERENCE: 2318-259

; CURRENT APPLICATION NUMBER: US/09/535,008

; CURRENT FILING DATE: 2000-03-23

; EARLIER APPLICATION NUMBER: U.S. 60/125,806

; EARLIER FILING DATE: 1999-03-23

; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 65

; LENGTH: 1679

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-535-008-65

Query Match 14.0%; Score 1008; DB 4; Length 1679;

Best Local Similarity 29.9%; Pred. No. 5.4e-75;

Matches 300; Conservative 158; Mismatches 325; Indels 222; Gaps 32;

Qy 269 LKGLLHPYQLEGLNFRFSWSKQTHVILADEMGLGKTIQSIALLASLFEENLI--PHLVI 326

Db 750 VNGVLKQYQIKGLEWLVSLYNNLNGILADEMGLGKTIQTIOTALITYLMEHCRINGPFLII 809

Qy 327 APLSTLRNWEREFATWAPQMVVYFGTAQARAVIREHEFYLSKDQKIKKKSGQISSE 386

Db 810 VPLSTLSNWAYEFDKWAPSVVYKGSPPARRAF-----VPQLRSG-----851

Qy 387 SKQRIKFDVLLTSYEMINLDSAVLKPIKWECEMIVDEGHRLLKNKDSKLFSSL-TQYSSNH 445

Db 852 -----KPNVLLTTYEYIIKDKHILAKIRWKYMIVDEGHRMKNHCKLTQVLNTHYVAPR 905

QY	446	RILLTGTPLONNDELPMMLMHFLDAGKFGSLEEFQBEFK-----DINOEEQ---I	492
DB	906	RLLLTGTPLOKLPDWALLNLLPTIFKSCSTPEQWFNAPFAMTGEKVLDNEEETILII	965
QY	493	SRLHKMLAPHLRRVKQDVMDMPKPKELILRVDLSSLQKEYY----KAIFTRNVQVLT	547
DB	966	RRLHKVLRPFLRLRLKKEVEAQLPEKVEYVIKCDMSALQRLVYRHMQAQGVLLTDGSEKD	1025
QY	548	KKG-GAQISLNNIMELRKVCCHPYMLEGVPEVIHDANEAFKQ-----LLES	593
DB	1026	KKGKGGTKTLMTNTINQLRKICNHPYMFQHIE-----ESPSEHLGFTGGIVQGLDLYRA	1078
QY	594	CGKLQLLDKMMVKLBQGHRVLIYTOFQHMULDLEDYCTHKKWQYERIDGKVGGAERQIR	653
DB	1079	SGKFELDRILPKLRATNHHKVLLFCQMTSLMTIMEDYFAYRGFKYLRDLGDTTKAEDRGML	1138
QY	654	IDRFNAKNSNKFCLLSTRAGGLGINLATADTVIYSDWNPHADLOAMARHRLGQTNK	713
DB	1139	LKTFNEPGSEYFIFLLSTRAGGLGLNLQSDADTVIIFSDWNPHQDLQAQDRAHRIGQQNE	1198
QY	714	VMIYRLINRGTIERRMMQLTKKMVLEHLVV--GKLKTQINQEE---LDDIIRYGSKE	768
DB	1199	VRVRLCTVNSVEEKILAAAKYKLVNDQKVIQAGMFDQKSSSHERRAFLQAILHE	1253
QY	769	FASEDDE-----AGKSGKIHY--DDAAIDKLLDRDLVEAEVSVDDDEENGFLKAFKV	819
DB	1254	--EODESRHCSTGSGSASFAHTAPPPAGVNPDLPEEPPLKEEDEVPDDETQNMIA	1310
QY	820	ANFEYIDENEAALAEQORVAES---KSSAGNSDRASYWEELLKDKFELH----QA	872
DB	1311	--EFDLFMRMDLDRREEARNPKRPRLMEEDLPSW--IIKDDAEVERLTCEEEEEK	1365
QY	873	ALGKRKRSRKQL-----VSIE	888
DB	1366	MFGGRGSRHRKEVDYSDLTEKQWLKKITGKDIDHTASSVARGLOFQRLQFCTRASKAIE	1425
QY	889	EDDLAGLEDV-----SSDGDES YEAEST-----DGEAAGQGVQTGRPPYRRKG	931
DB	1426	EGTLEEIEEEVRQKSSRRKRKRDSDAGSSTPTTSTRSRDKDDESKQ-----KKRG	1476
QY	932	R---DNLEPTP-----LMEGEGRSFRVLGFNQSQRAIFVQTLMRVYGAGNF-	973
DB	1477	RPPAEKLSNPPNLTCKMKKIVDAVIKYKDSSSGRQLS-----EVFIQLPSRKELPEY	1530
QY	974	-----DWKEFVPRLKQKTFEEINEYGILFLKHTAAEIDENSPTFSDGVPKEG-LRIED	1025
DB	1531	ELIRKPVDFFKKIERIRNHKYRSIND-----LEKDVMLLCQNAQTFN---LE	1581
QY	1026	VLVRIALLILVQEKVKFVEDHPGKPVFPSPRILERFPGLRSGKIWKEEHDKIMIRAVLKHG	1085
DB	1582	SIVLQSVFTSVRQKIEKEDDSEGE-----ESEEEEEEGEEGSESESRSVKV	1630
QY	1086	YGRWQAIVDCKELGIELICKELNFPHISLSAAEQAGLQQONGSG	1130
DB	1631	LGRKEKAQDRLKGRRRPRSGSRAPKPVVSDDDSEEEEOEDRS	1675

## RESULT 13

US-09-702-705-1819  
; Sequence 1819, Application US/09702705  
; Patent No. 6504010  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Bangur, Chaitanya S.  
; APPLICANT: Lodes, Michael A.  
; APPLICANT: Fanger, Gary  
; APPLICANT: Vedvick, Tom  
; APPLICANT: Carter, Darrick  
; APPLICANT: Retter, Marc  
; APPLICANT: Mannion, Jane  
; APPLICANT: Fan, Liqun  
; TITLE OF INVENTION: COMPOSITIONS AND  
; TITLE OF INVENTION: DIAGNOSIS OF LUNG

```

; FILE REFERENCE: 210121.478C14
; CURRENT APPLICATION NUMBER: US/09/702,705
; CURRENT FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 1833
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1819
; LENGTH: 831
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-702-705-1819

Query Match      13.9%; Score 997; DB 4; Length 831;
Best Local Similarity 36.3%; Pred. No. 1.5e-74;
Matches 246; Conservative 114; Mismatches 212; Indels 106; Gaps 19

QY 257 RDFQQFDHTPEFLKGL-LHPYQLEGLNFLFRFSWSKQTHVILADEMGLGKTIQSIALL--- 312
Db 34 QDLRWG-----LTGIHLRSYQLEGVNWLQRFHCHQNCILGDEMGLGKTCQTIALFIYL 88

QY 313 -ASLFEENLPHLVIAPLSTLRNWEREFATWAPQMVVYFGTAQARAVIREHEFYLSKD 371
Db 89 AGRLNDEG--PFLICPLSVLSNWKEEMQRFAPGLSCVYAGDKEERACLOQ----- 138

QY 372 QKIKKKKSGQISSESQKRIKFDVLLTSYEMINLDSAVLKPWKECMIVDEGHRLLNKD 431
Db 139 -----DLKQES----RFHVLLTTYEICLKDASFLKSPFWSVLVWDEAHLRKNQS 183

QY 432 SKLFSSLTQYSSNHRILLTGTPLONNLDELFMLMHFLDAGKFGSLE--EFQEEFKDINQE 489
Db 184 SLLHKTLSEFSVVSFLLLTGTPIONSLQELYSLLSFVEPDLFSKEEVGDFIORYQDIEKE 243

QY 490 -EQISRLHKMLAPHLLRRVKDVMKMPKKELILRVDLSSLQKEYYKAIFTRNYQVLT 548
Db 244 SESASELHKLLQPLLRVRVKAEVATELPKKTTEVVYHGMSALQKKYKAILMKDLDLAFEN 303

QY 549 KGAQISLNNIMELRKVCCHPYMLEGVPEVIHDANEAFKQLLESCGLQLLDKMMVKLK 608
Db 304 ETAKVKVLQNILSQLRKCVDPHPYLPFDGVEP---EPFEVGDHLTEASGKLHLLDKLLAFLY 360

QY 609 EQGHRVLIYTFQHMULDLEDYCTHKKWQYERIDGKVGGAERQIRIDRFNAKNSNKFCFL 668
Db 361 SGGHRVLLFSQMTQMDILQDYMRYGYSYERVDGSRGEERHLAIKNFGQQPI--FVFL 418

QY 669 LSTRAGGLGINLATADTVIIYDSDNWPHADLOAMARAHLGQTNKVMYIRLINRGTIEE- 727
Db 419 LSTRAGGVGMNLTAADTVIFVDSDFNPQNDLQAAARAHRIGQNKSVKIRLIGRDTVEEI 478

QY 728 -----RVMQLTKKKQVLE--HLVVGKLKTONINQEBLDDIIRYGSKELFASEDDEAGKS 779
Db 479 VYRKAASKLQLT--NMIEGGHFTLGAQKPAADADLQSEILKFLGLDKLLASEGST----- 532

QY 780 GKIHYDDAAIDKLDR-----DLVEAEVSVDDDEENGFLKAFKVANFEYIDENEA 831
Db 533 ---MDEIDLSEILGETKDGQWVSDALPAEAGSGSRDQEG-----KNHMYLFEKGDY 580

QY 832 ALEAQRVAAESKSSAGNSDRASYWEELLKDKFELHQAEE-----LNALGKRKR 879
Db 581 SKEPSK-----EDRKSFEQLVNLQTLTLLKASQEGRSLRNGKSVLIPGLVEGST 629

QY 880 SRKQLVSIEEDDLAGLED 897
Db 630 KRKRVLSPPE-----LED 642

```

RESULT 14  
US-09-736-457-1819  
; Sequence 1819, Application US/09736457  
; Patent No. 6509448  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Bangur, Chaitanya S.  
; APPLICANT: Lodes, Michael A.  
; APPLICANT: Farger, Gary

```

; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; APPLICANT: Wang, Aijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1819
; LENGTH: 831
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-736-457-1819

Query Match      13.9%; Score 997; DB 4; Length 831;
Best Local Similarity 36.3%; Pred. No. 1.5e-74;
Matches 246; Conservative 114; Mismatches 212; Indels 106; Gaps 19;

Qy 257 RDFQQFDHTPEFLKGL-LHPYQLEGLNFLRFSWSKQTHVILADEMGLGKTIQSIALL--- 312
Db 34 QDLRWG-----LTGIHLRSYQLEGVNWLAQRFHCQNGCILGDEMGLGKTCQTIALFIYL 88

Qy 313 -ASLFEENLIPHLVIAPLSTLRNWEREFATWAPQMNVMYFGTAQARAVIREHEFYLSKD 371
Db 89 AGRLNDEG--PFLILCPLSVLSNWKEEMQRFAPGLSCVTYAGDKERACLQ----- 138

Qy 372 QKKIKKKSGQISSESQKRIKFDVLLTSYEMINLDSAVLKPIKWECMIVDEGHRLLKNKD 431
Db 139 -----DLKQES-----RFHVLLTYEICLKDASFLLKSFPSVSVVDEAHRLLKNQS 183

Qy 432 SKLFFSSLTQYSSNHRILLTGTPLNNDLDELFLMHFLDAGKFGSLE--EFQEEFKDINQE 489
Db 184 SLHKTLSEFSVVFSLLLTGTPIONSLQELYSLLSFVEPDLFSKEEVGDFIORYQDIEKE 243

Qy 490 -EQISRLHKMLAPHLRRVKQDVMDMPKPKKELILRVDLSSLOKEYYKAIFTRNYQVLT 548
Db 244 SESASELHKLLQPFLLRRVKAEVATELPKKTTEVVIYHGMSALQKKYKAILMKOLDAFEN 303

Qy 549 KGAQISLNNIMMELRKVCCHPYMLEGVPEVIHDANEAFKQLLES CGKLQLLDDKMMVKLK 608
Db 304 ETAKKVKLQNILSQLRKCVDPHPYLFDFGVEP--EPFEVGDHLTEASGKLHLLDKLLAF 360

Qy 609 EQGHRVLIYTFQHMLDLLEDYCTHKKQYERIDGKVGGAERQIRIDRFNAKNSKFCFL 668
Db 361 SGGHRVLLFSQMTQMLDILQYMDYRGYSYERVDGSGVRGEERHLAIKNFGQQPI--FVFL 418

Qy 669 LSTRAGGLGINLATADTVIIYSDSNPHADLOAMARAHRGLQTNKVMYIRLINRGITIEE- 727
Db 419 LSTRAGGVGMNLTAADTVIFVDSDFNPQNDLQAAARAHRIGQNKSVKVIIRLIGRDTVEEI 478

Qy 728 -----RMMQLTKKQWLE--HLVVGKLTQNIQNEELDDIIRYGSKELFASEDDEAGKS 779
Db 479 VYRKAASKLQLT--NMIEGGHFTLGAQKPAADADLQLEILKFLGDLKLASEGST----- 532

Qy 780 GKIHYDDAAIDKLLDR-----DLVEAEVSVDDDEENGFLKAFKVANFEYIDENEAA 831
Db 533 ----MDEIDLLESILGETKDGQWVSDALPAAEGGSRDQEEG-----KNHMYLFEKG 580

Qy 832 ALAQORVAESKSSAGNSDRASYWEELLKDKFELHQAEE-----LNALGKRKR 879
Db 581 SKEPSK-----EDRKSFEQLVNLQTLLEKASQEGSLRNKGSVLIPGLVEGST 629

Qy 880 SRKQLVSIIEEDDLAGLED 897
Db 630 KRKRVLSPPE-----LED 642
```

```

US-09-671-325-1819
; Sequence 1819, Application US/09671325
; Patent No. 6667154
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C12
; CURRENT APPLICATION NUMBER: US/09/671,325
; CURRENT FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 1825
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1819
; LENGTH: 831
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-671-325-1819

Query Match      13.9%; Score 997; DB 4; Length 831;
Best Local Similarity 36.3%; Pred. No. 1.5e-74;
Matches 246; Conservative 114; Mismatches 212; Indels 106; Gaps 19;

Qy 257 RDFQQFDHTPEFLKGL-LHPYQLEGLNFLRFSWSKQTHVILADEMGLGKTIQSIALL--- 312
Db 34 QDLRWG-----LTGIHLRSYQLEGVNWLAQRFHCQNGCILGDEMGLGKTCQTIALFIYL 88

Qy 313 -ASLFEENLIPHLVIAPLSTLRNWEREFATWAPQMNVMYFGTAQARAVIREHEFYLSKD 371
Db 89 AGRLNDEG--PFLILCPLSVLSNWKEEMQRFAPGLSCVTYAGDKERACLQ----- 138

Qy 372 QKKIKKKSGQISSESQKRIKFDVLLTSYEMINLDSAVLKPIKWECMIVDEGHRLLKNKD 431
Db 139 -----DLKQES-----RFHVLLTYEICLKDASFLLKSFPSVSVVDEAHRLLKNQS 183

Qy 432 SKLFFSSLTQYSSNHRILLTGTPLNNDLDELFLMHFLDAGKFGSLE--EFQEEFKDINQE 489
Db 184 SLHKTLSEFSVVFSLLLTGTPIONSLQELYSLLSFVEPDLFSKEEVGDFIORYQDIEKE 243

Qy 490 -EQISRLHKMLAPHLRRVKQDVMDMPKPKKELILRVDLSSLOKEYYKAIFTRNYQVLT 548
Db 244 SESASELHKLLQPFLLRRVKAEVATELPKKTTEVVIYHGMSALQKKYKAILMKOLDAFEN 303

Qy 549 KGAQISLNNIMMELRKVCCHPYMLEGVPEVIHDANEAFKQLLES CGKLQLLDDKMMVKLK 608
Db 304 ETAKKVKLQNILSQLRKCVDPHPYLFDFGVEP--EPFEVGDHLTEASGKLHLLDKLLAF 360

Qy 609 EQGHRVLIYTFQHMLDLLEDYCTHKKQYERIDGKVGGAERQIRIDRFNAKNSKFCFL 668
Db 361 SGGHRVLLFSQMTQMLDILQYMDYRGYSYERVDGSGVRGEERHLAIKNFGQQPI--FVFL 418

Qy 669 LSTRAGGLGINLATADTVIIYSDSNPHADLOAMARAHRGLQTNKVMYIRLINRGITIEE- 727
Db 419 LSTRAGGVGMNLTAADTVIFVDSDFNPQNDLQAAARAHRIGQNKSVKVIIRLIGRDTVEEI 478

Qy 728 -----RMMQLTKKQWLE--HLVVGKLTQNIQNEELDDIIRYGSKELFASEDDEAGKS 779
Db 479 VYRKAASKLQLT--NMIEGGHFTLGAQKPAADADLQLEILKFLGDLKLASEGST----- 532

Qy 780 GKIHYDDAAIDKLLDR-----DLVEAEVSVDDDEENGFLKAFKVANFEYIDENEAA 831
Db 533 ----MDEIDLLESILGETKDGQWVSDALPAAEGGSRDQEEG-----KNHMYLFEKG 580

Qy 832 ALAQORVAESKSSAGNSDRASYWEELLKDKFELHQAEE-----LNALGKRKR 879
Db 581 SKEPSK-----EDRKSFEQLVNLQTLLEKASQEGSLRNKGSVLIPGLVEGST 629
```



Qy 880 SRKQLVSI EEDDLA GLED 897  
| | : : | | | | | |  
Db 630 KKRVLSP EE-----LED 642

Search completed: September 14, 2004, 02:09:21  
Job time : 44 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 14, 2004, 02:05:02 ; Search time 88 Seconds  
(without alignments)  
5043.573 Million cell updates/sec

Title: US-10-049-137-2  
 Perfect score: 7187  
 Sequence: 1 MSSLVRLRLIRSDRKPVYNL.....VDVXMEEAKEEEKPKNMVVD 1384

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1335176 seqs, 320689617 residues

Total number of hits satisfying chosen parameters: 1335176

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```

Database : Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

```

## SUMMARIES

Result No.	%			DB	ID	Description
	Score	Match	Length			
1	3752	52.2	1358	16	US-10-675-072A-43	Sequence 43, Appl
2	3621	50.4	1364	16	US-10-437-963-155910	Sequence 155910,
3	2105.5	29.3	1954	12	US-10-462-261-2	Sequence 2, Appl
4	2083	29.0	1944	16	US-10-408-765A-545	Sequence 545, App
5	1874.5	26.1	979	15	US-10-104-047-2446	Sequence 2446, Ap
6	1621.5	22.6	1739	15	US-10-341-434-61	Sequence 61, Appl
7	1570.5	21.9	1795	8	US-08-973-363-17	Sequence 17, Appl
8	1555	21.6	2108	8	US-08-973-363-16	Sequence 16, Appl
9	1551.5	21.6	1434	8	US-08-973-363-20	Sequence 20, Appl
10	1548.5	21.5	2192	16	US-10-437-963-201156	Sequence 201156,
11	1536.5	21.4	1465	15	US-10-369-493-5100	Sequence 5100, Ap
12	1492.5	20.8	1967	10	US-09-849-602-16	Sequence 16, Appl
13	1473	20.5	1407	15	US-10-369-493-2310	Sequence 2310, Ap
14	1461.5	20.3	1467	8	US-08-973-363-21	Sequence 21, Appl
15	1461.5	20.3	1468	15	US-10-369-493-1775	Sequence 1775, Ap

16	1426.5	19.8	1264	15	US-10-369-493-3742	Sequence 3742, Ap
17	1407.5	19.6	2131	16	US-10-408-765A-1179	Sequence 1179, Ap
18	1366	19.0	1413	16	US-10-437-963-138262	Sequence 138262, A
19	1356	18.9	1061	15	US-10-369-493-12847	Sequence 12847, A
20	1293.5	18.0	1141	16	US-10-437-963-106873	Sequence 106873, A
21	1241	17.3	1152	12	US-10-412-699B-1581	Sequence 1581, Ap
22	1241	17.3	1158	16	US-10-437-963-154461	Sequence 154461, A
23	1230	17.1	1104	12	US-10-425-114-57274	Sequence 57274, A
24	1224.5	17.0	1057	12	US-09-934-455-432	Sequence 432, App
25	1224.5	17.0	1057	12	US-10-412-699B-790	Sequence 790, App
26	1224.5	17.0	1057	12	US-10-412-699B-1846	Sequence 1846, Ap
27	1224.5	17.0	1057	12	US-10-225-066A-1044	Sequence 1044, Ap
28	1224.5	17.0	1057	15	US-10-374-780A-2554	Sequence 2554, Ap
29	1215	16.9	1122	12	US-10-412-699B-1582	Sequence 1582, Ap
30	1215	16.9	1122	16	US-10-437-963-151729	Sequence 151729, A
31	1157	16.1	730	16	US-10-408-765A-1039	Sequence 1039, Ap
32	1147	16.0	659	12	US-10-424-599-226981	Sequence 226981, A
33	1142	15.9	620	12	US-10-425-114-37386	Sequence 37386, A
34	1139	15.8	1120	15	US-10-369-493-22416	Sequence 22416, A
35	1137.5	15.8	956	15	US-10-369-493-3306	Sequence 3306, Ap
36	1135.5	15.8	1143	15	US-10-369-493-1499	Sequence 1499, Ap
37	1131	15.7	1027	14	US-10-205-219-123	Sequence 123, App
38	1126	15.7	971	15	US-10-369-493-5755	Sequence 5755, Ap
39	1120	15.6	1359	15	US-10-369-493-22090	Sequence 22090, A
40	1097	15.3	1302	14	US-10-032-585-7534	Sequence 7534, Ap
41	1095.5	15.2	1703	9	US-09-801-368-340	Sequence 340, App
42	1095.5	15.2	1703	10	US-09-824-574-3	Sequence 3, Appli
43	1095.5	15.2	1703	15	US-10-369-493-22413	Sequence 22413, A
44	1093.5	15.2	1474	15	US-10-369-493-6164	Sequence 6164, Ap
45	1093.5	15.2	1474	15	US-10-369-493-6165	Sequence 6165, Ap

## ALIGNMENTS

## RESULT 1

```

US-10-675-072A-43
; Sequence 43, Application US/10675072A
; Publication No. US20040098760A1
; GENERAL INFORMATION:
; APPLICANT: Yumin, Tao
; APPLICANT: Gordon-Kamm, William
; APPLICANT: Shen, Bo
; APPLICANT: Lowe, Keith
; APPLICANT: Danilevskaya, Olga
; APPLICANT: Mahajan, Pramod
; APPLICANT: Rafalski, Jan Antoni
; APPLICANT: Sakai, Hajime
; APPLICANT: Klein, Theodore
; TITLE OF INVENTION: Transcriptional Regulation of
; TITLE OF INVENTION: Acids, Polypeptides,
; FILE REFERENCE: 1288R
; CURRENT APPLICATION NUMBER: US/10/675,072A
; CURRENT FILING DATE: 2003-09-30
; PRIOR APPLICATION NUMBER: 10/005,057
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/251,555
; PRIOR FILING DATE: 2000-12-06
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 1358
; TYPE: PRT
; ORGANISM: Zea mays
US-10-675-072A-43

```

Query Match	52.2%;	Score 3752;	DB 16;	Length 1358;
Best Local Similarity	54.7%;	Pred. No. 1.5e-272;		
Matches 765:	Conservative 227;	Mismatches 305;	Indels 102;	Gaps 32;

```
QY      1 MSSLVERLIRSDRKPVYNLDSDDDDFVP-----KKORTFEQVEAIVRTDAKENAC 52
        |||||:||:||:||:||:||:||:||:||:||:||:||:||:||:||:||:||
        |||||:||:||:||:||:||:||:||:||:||:||:||:||:||:||:||:||

pb      1 MSSLVERLVRSERRPRVALDESDDD--LPLRVGAGKGKQQNDAPAEIRTEREAKKEAAC 58
```





Db 253 NKRVIILGDEMGLGKTIQSI AFLGSLFVDKLGPHLVVAPLSTLRNWEREFATWAPQMNVM 312  
QY 351 YFGTAQARAVIREHEFYLSKDQ-KKIKKKSGQISSSESKQRIKFDVLLTSYEMINLDSA 409  
Db 313 YFGSAASREIIRKIEFYYPKEPKKLLKKKSSPSNEDKKQSRKIKFDVLLTSYEMINMDST 372  
QY 410 VLKPIKWECMIVDEGHRLLKNKOSKLFSSLTQYSSNHRILLTGTPQNNDLFLMLMHFLD 469  
Db 373 VLKTIIEWECMIVDEGHRLLKNKOSKLFQGLKBYHTKHRVLLTGTPVQNNDLFLMLMHFLE 432  
QY 470 AGKFGSLEEFQEEFKDINQEEQISRLHKMLAPHLLRRVKOVKMDMPKKELILRVDLSS 529  
Db 433 GDSFGSIADLQEEFKDINQDKQVEKLHGMLKPHLLRRFKOVKMKELPPKKELILRVELTS 492  
QY 530 LQKEYYKAIFRNYQVLTKKGGAQISLNNIMMELRKVCCHPYML-EGVEPV----- 579  
Db 493 KQKEYYKAILTKNYEVLTRRSQGHVSLINVVMELRKLCHAFMTDEPEEPANSEEARFS 552  
QY 580 -----IHDANEAPK-----QLLESCGK 596  
Db 553 DVINILPLPAPYFQTHIDAEVLEFSDISCCYLPISRLNKHNAIFEVVDIWRLLESSGK 612  
QY 597 LQLLDKMVKLKEQGHRVLIYTQFQHMLDLLEDYCTHKKWQYERIDGKVGAERQIRIDR 656  
Db 613 MELLDKMVKLKEQGHRVLIYSQFQHMLDLLEDYLSYRKWSYERIDGKIGGAERQIRIDR 672  
QY 657 FNAKSNKFCFLSTRAGGLGINLATADTVIIYSDWNPHADLQAMARAHRLGQTNKVM 716  
Db 673 FNAKNSTRFCFLSTRAGGLGINLATADTVIIYSDWNPHADLQAMARAHRLGQTSKVM 732  
QY 717 YRLINRGTTIEERMQLTKKMWLEHLVVGKL-KTONINQEEELDDIIRYGSKELFASEDDE 775  
Db 733 YRLVSRGTTIEERMQLTKKMWLEHLVVGRLTKGTNIVQEEELDDIIRHGSKELFDDENDE 792  
QY 776 AGKSGKIHVDDAAIDKLLDRDLVEAEESVSDDEENGFLKAFKVANFEYIDENEAALAE 835  
Db 793 AGKSCQIHVDDAAIDRLDRDQADGEE-PVEDEEEDEFKGFKVANFEYIDEAKALAAKE 851  
QY 836 QRVAAESKSAGNSDRASYWEELLKDKPELHQABEELNALGKRKRKQLVSIIEEDDLAGL 895  
Db 852 EEARKKAEBEAANSDRANFWDKLDRYDVQKVEEHTTMGKGRKRQMAAADEDDITGL 911  
QY 896 EDVSS-DGDESYEAESTDGEAAGQGVQTGRR-PY-RRKGRDNLEPTPLMEGEGRSFRVLG 952  
Db 912 HDMSSDEDDYSYDDDDVSDNDTSLQSLAGRRGPYSKKQRSNVDSLPPFMEGEGRALRVYG 971  
QY 953 ENQSQRATFVOTLMRYGAGNFDWKEFVPRLKQKTFEEINEYGILFLKHIAEEIDENSPTF 1012  
Db 972 FNQIORTQFOTLMRYGFQNYDWKEFTPLRGKSVEEIQRYAELVMTHLLEDIND-SGY 1030  
QY 1013 SDGVPKEGLRIEDVLRIALLILVQEKVFEVDHPGKPVFPSPRIERFPGLRSGKIWKEE 1072  
Db 1031 ADGVPKE-MRTDETIVRLANISLVEEKVAAMEQGKITKLFPSYLLYEFPSLVGGRVWKA 1089  
QY 1073 HDKIMIRAVLKHGYGRWQAIIVDDKELGIQELICKELNFPHISLSAABQAGLQONGSGGS 1132  
Db 1090 QDLLLLKALIKHGYARWQYISDDRDNNGIFEAAARQELRP-----TANELISSHSNN----- 1140  
QY 1133 NPGAQTNQNPQSVITGNNNASADGAQVNSMFYRDMQRRLLVEFKVRVLLLEKAMNYEYA 1192  
Db 1141 ----ETNGNLESTQEQSNPT-----SMIHYRDTQRKIVEFIRKRYHLLERCLNLEYA 1189  
QY 1193 EYYGGLGSSSIPTPEEPAE-----PKIADTVGVSVFIEVDDEMDGLPKTDPITSEEIM 1246  
Db 1190 V----IKTKTPVDDDLAEQDFPGGHRPAVPDY-----SEMRELVPLEPISKEVAP 1236  
QY 1247 GAAVDNNQARVEIAQHYNQMCKLLDENARESVOAYVNNQPPSTKYNESFRALKSINGNIN 1306  
Db 1237 EGTDDQSQ-----VSHLYNKMCFVLEDSAVPALNSHFQDKAASSGLANSLHKFEAVCEDVS 1292  
QY 1307 TILSITSDQKSHEDDTKPDNNVEMKDTABETKPLRGGVVDLNVVEGEENIAEASGSVD 1366

Db 1293 RIL-----RSHENGTTPK-EEVMLDASSKETTSKPDPATEVPSASK-----EATPPVQ 1340  
QY 1367 VKMEEAKEEEKP 1378  
Db 1341 DPVIEAVKKEPP 1352  
RESULT 3  
US-10-462-261-2  
; Sequence 2, Application US/104622261  
; Publication No. US20040029248A1  
; GENERAL INFORMATION:  
; APPLICANT: Garrett M. Brodeur  
; APPLICANT: Peter S. White  
; TITLE OF INVENTION: CHD5 ENCODING NUCLEIC ACIDS,  
; TITLE OF INVENTION: POLYPEPTIDES, ANTIBODIES AND METHODS OF USE THEREOF  
; FILE REFERENCE: CHOP0162  
; CURRENT APPLICATION NUMBER: US/10/462,261  
; CURRENT FILING DATE: 2003-06-16  
; PRIOR APPLICATION NUMBER: 60/388,848  
; PRIOR FILING DATE: 2002-06-14  
; NUMBER OF SEQ ID NOS: 69  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1954  
; TYPE: PRT  
; ORGANISM: Homo Sapiens  
US-10-462-261-2

Query Match 29.3%; Score 2105.5; DB 12; Length 1954;  
Best Local Similarity 35.9%; Pred. No. 2e-148;  
Matches 533; Conservative 227; Mismatches 437; Indels 289; Gaps 49;

QY 21 DDSDDDDFVPKKDRTFEQVEAIVRTDAKENACQACGESTNLVSCNTCTYAFHAKCLVPPL 80  
Db 395 EPKDDDD--EEEGGCCEEE-----DDHMEFCRVCKDGGELCCDACPSSYHLHCLNPPL 447  
QY 81 KDASVENWRCPECVSP--LNEIDKILDCE-MRPTKSSEQ--SSDAB-----PKP---IF 127  
Db 448 PEIPNGEWLCPRCTCPPPLKKGQVRIHLHWRWTEPPAPFMVGLPGPDVEPSPPLPPKLEGIP 507  
QY 128 VKQYLVKWKGLSYLHCSWVPEKEFQKAYKSNHRLKTRVNNFHR-----QMES 174  
Db 508 EREFFVKWAGLSYWHCSWVKELQLELYHTVMYRNYQRKNDMDPPPPDFYSGSDEDEGKSEK 567  
QY 175 FNNS-----EDDF--VAIRPEWTTVDRIIL-ACREEDGELEYLVKYKELSYDECYWE-S 223  
Db 568 RKNKDPYAKMEERFYRYGIKPEWMMIHRILNHSFDKGDVHYLIKWKDLPYDQCTWEID 627  
QY 224 ESDISIFQNEIQRF-----KDVNSRTRRSKDVHDKRNPRDFQFQDHTPEF 268  
Db 628 DIDIPYDNLKQAYWGHRELMLGEDTRLPKRLLLKGGKLLRD-DKQEKPPDTPIVDPTVKF 686  
QY 269 LK-----GLLHPYQLEGNLFRFSWSKQTHVILADEMGLGKTIQSIALLASLFEE- 318  
Db 687 DKQPMYIDSTGGTLHPYQLEGNLWLRFSWAQGTDTILADEMGLGKTVQTIIVFLYSLYKEG 746  
QY 319 -NLIPLHVIAPLSTLRNWEREFATWAPQMNVMVYFGTAQARAVIREHEFYLSKDQKIKK 377  
Db 747 HSKGPYLVSAPLSTIINWEREFEMWAPDFYVVTYTGDKESRSVIRENEF--SPEDNAIRS 804  
QY 378 KKSQGISSESQKRIKFDVLLTSYEMINLDSAVLKPIKWECMIVDEGHRLLKNKOSKLFSS 437  
Db 805 GK--KVPFMKKEVQIKFHVLLTSYELITIDQAILGSIEWACLIVDEAHLKNNQSKFFRV 862  
QY 438 LTQYSSNHRILLTGTPQNNDLFLMLMHFLDAGKFGSLEEFQEEFKDINQEEQISRLHK 497  
Db 863 LNSYKIDYKLLLTGTPLQNNLEELFHLNFLTPTPERFNNLEGLEFEEFADISKEDQIKKLDH 922  
QY 498 MLAPHLLRRVKKDVMDKMPKKELILRVDLSSSQKEYYKAIFRNYQVLTKKGGA-QISL 556  
Db 923 LLGPHMLRRLKADVFNKMPAKTELIIVRVELSQMKYKYKFILTRNFEALNSKGGGNQVSL 982











; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,363
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/GB96/01341
; FILING DATE: 05-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9511439.3
; FILING DATE: 06-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER: 263/PPNTIR1172US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)-721-8200
; TELEFAX: (202)-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2108 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-973-363-16

Query Match 21.6%; Score 1555; DB 8; Length 2108;
Best Local Similarity 30.6%; Pred. No. 5.9e-107;
Matches 430; Conservative 242; Mismatches 411; Indels 320; Gaps 49;
QY 7 RLRIIRSDRPV-----YNLDDS-----DDDFVPKKDR-----TFEQVEAIVRTDAKEN 50
Db 245 RIKPSGKKGSTGQKKRQLDSDSEEDDDDEDYDKGRSRRQATVNVSYKEAEE-TKTDSD-DD 302
QY 51 ACQACGESTNLVSCNTCTYAFHAKCLVPPLKDASVENWRCPECVSPLENIDKILDCEMRP 110
Db 303 LLEVCGED-----VPQTEDEFET-----IEKFMD--SRI 330
QY 111 TKSEQSGSS-----DAEPKPIFVK-----QYLVKWKGLSYLHCSVWVPEKEFQKAY 155
Db 331 GRKGATGASTTIYAVEADGDPNAGFEKSKELGEIQYLIKWKGWSHINTWETEETLK--- 387
QY 156 KSNHRLKTRVNNFHRQ-----MESFNNSEDDFVAIRPEWTTVDRIILACRE 200
Db 388 QONVKGMMKLDNYKKKQDKQETRWLKNASPEDVEYVNCQQLTDDLHKQYQIVERIIAHSN 447
QY 201 ED---GELEYLVKYEKELSYDECYWESESDIS-TFQNEIQRFKQVNV-SRTRRSKQVD-HKR 254
Db 448 QKSAAGYPDYCKWQGLPYSECSWEDGALIAKFKQARIDEYFSRNQSKTTPFKDCKVLKQ 507
QY 255 NPRDFQQFDHTPEFLKG----LLHPYQLEGNLFRLFSWSKQTHVILADEMGLGKTIQISIA 310
Db 508 RPR-FVALKKQPSYIGGHESLELRDYQLNGLNWLHAHSWCKGNSCILADEMGLGKTIQTIS 566
QY 311 LLASLFEENLI--PHLVIAPLSTLRNWEREFATWAPQMNVMYFGTAQARAVIREHEFYL 368
Db 567 FLNYLFHEHQLYGFLLRVPLSTLTWSQREIQTWAPQMNNAVYLGDITSRNWIRTHEW-- 624
QY 369 SKDQKKIKKKSGQISSKSKQRIKFDVLLTSYEMINLDSAVLKPIKWCMIWDEGHRLLK 428
Db 625 -----MHPQTKRLKFNILLTYEILLKDKSFLGGLNWAFIGVDEAHRLLK 668
QY 429 NKDSKLFSSLTQYSSNHRILLTGTPQLQNNLDELFMLMHFLDAGKFGSLEEFQEEFKDINQ 488
Db 669 NDDSLLYRTLIDFKSNHRLLLITGTPLQNSLKELWSLLHFIMPEKFFSSWEDFEEHKG-GR 727
QY 489 EEQISRLHKMLAPHLLRRVKKQVMKOMPPEKELILRVDLSLQKEYYKAIIFTRNYQVLTK 548
Db 728 EYGYASLHKELEPFLRRVKKQVKSLEPAKVEQILRMEMSALQKQYKWLITRNYKALSK 787

QY 549 -KGAQISLNNIMMELRKVCCHPYMLEGVEP-VIHDANEAFKQLLESCGKLQLLDKMMVK 606
Db 788 GSKGSTSGFLNIMMELKCCNHCYLIKPPDDNEFYNKQELQHLIRSSGKLILLDKLLIR 847
QY 607 LKEQHRVLIYTFQFHMLDLEEDYCTHKKQWYERIDGKVGGAERQIRIDRFNAKNSNKFC 666
Db 848 LRERGNRVLIIFSQVRMLDILAEYLYKYRQFPFQRLDGSIKELRKQALDHFNAEGSEDFC 907
QY 667 FLLSTRAGGLGINLATADTVIYSDWNPHADLOAMARHRLGQTNKVMYIRLINRGITIE 726
Db 908 FLLSTRAGGLGINLASADTVVIFSDWNPNQNDLQAAQARAHRIQKKQVNIYRLVTKGSVE 967
QY 727 ERMQLTKKKMVLHLVVGKLTQN-----INQBEELDDIIRYGSKELFAS 771
Db 968 EDILERAKKKMWLDHLVIQRMDDTGKTVLHTGSAPSSSTPFNKEELSAILKFGAEELFKE 1027
QY 772 EDDEACKSGKIHYDDAAIDKLLDRDLVEAEE--VSVDDEENGFLKAFKVANFEYIDEN 828
Db 1028 PEGEEQEP-----QEMDIDEILKRAETHENEPGLPSVGDE---LLSQFKVANFSNMDED 1078
QY 829 EAAALEAQRVAEASKSSAGNSDRASYWEEL-----LXDKFELHQAEEELNALGKRKRS 880
Db 1079 D-IELEPER---NSKN-----WEEIIPEEQRRRLEEEERQKELEEIYMLPRMNC 1124
QY 881 RKQLVSIIEEDDLAGLEDVSSGDSEYAEASTDGEAAGQGVQTCRRPYRRKGRDNLEPTPL 940
Db 1125 AKQ-----ISFNGSEGRRSRRRYSGSDSDSISERKKRPKKRGRPTIPRE- 1169
QY 941 MEGEGRSFRVLGFNQSORAIFVQTLRMRYGAGNFDWKEFVR---LKQKTFEEINEYGILF 997
Db 1170 -----NIKGFSDAEIRRFIKSYKKFG-GPLERLDAIARDAELVDKSETDLRRLGELV 1220
QY 998 LKHIAEEIDENS-----PTFSDGVPKEGLRIEDVLVRIALLILVQEKVKF 1042
Db 1221 HNGCVKALKDSSSGTERAGRLGKVKGPFTF-----RISGVQVNAKLVIAHEDEL-- 1269
QY 1043 VEDHPGKPVFPSPRILERFPGLRSKI-----WKEEHDKIMIRAVLKHGYGRWQ 1090
Db 1270 IPLHKSIPSDPE---ER---KQYTIPCHTKAAHFDIDWGKEDDSNLLIGIYEGYGSWE 1322
QY 1091 AIVDDKELGIGQELI-----CKELNFPH---ISLSAAEQAGLQGN--GSGGSNPG 1135
Db 1323 MIKMPDLSLTHKILPDDPKKPKQAKQLQTRADYLIKLSRDLAKREAQRLCAGGS--- 1379
QY 1136 AQTNQNPGSVITGNNASADGAQVNSMFYRDMQRRVFEVFKRVLLLEKAMNVEYAEY 1195
Db 1380 -----KRRKTRAKSKSAM-----KSIKVKKEI 1401
QY 1196 YGLGSSSIPTEEPEAEPIADI TVGSVFIEVDDEMLDGLPKTDPTITSEEIM----- 1246
Db 1402 --KSDSSPLPSEKSD-----EDDDKLNDSKPESKDRSKSVSDAPVHITA 1445
QY 1247 -GAAVDNNQARVEIAQHYNQMCK 1268
Db 1446 SGEPVPIAESEELDQKTFISICK 1468

RESULT 9
US-08-973-363-20
; Sequence 20, Application US/08973363
; Publication No. US20030191297A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: AVIAN GHD GENES AND THEIR USE IN METHODS FOR
; TITLE OF INVENTION: SEX IDENTIFICATION IN BIRDS
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WENDEROTH, LIND & PONACK, L.L.P.
; STREET: 2033 K. Street, N.W., Suite 800,
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.



ZIP: 20006  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/973,363  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/GB96/01341  
FILING DATE: 05-JUN-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9511439.3  
FILING DATE: 06-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren M Cheek, Jr.  
REGISTRATION NUMBER: 33,367  
REFERENCE/DOCKET NUMBER: 263/PPNTR1172US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)-721-8200  
TELEFAX: (202)-721-8250  
TELEX:  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1434 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..1434  
OTHER INFORMATION: /note= "The sequence beginning at 1  
corresponds to 55 and that ending at 1434 corresponds to  
1488"  
OTHER INFORMATION: 1488"  
US-08-973-363-20

Query Match 21.6%; Score 1551.5; DB 8; Length 1434;  
Best Local Similarity 32.4%; Pred. No. 5.8e-107;  
Matches 408; Conservative 220; Mismatches 389; Indels 243; Gaps 40;  
Qy 7 RLIRSDRPV-----YNLDS-----DDDFVPKKDR-----TFEQVEAIVRTDAKEN 50  
Db 117 RIKPSGKXSTGQKKRQLDSSEEEEDDEDYDKGSRRTQATVNVSYKEABE-TKTDSD- 174  
Qy 51 ACQACGESTNLVSCNTCTYAFHAKCLVPPLKDAVENWRCPECVSPLEIDKILDCMRP 110  
Db 175 LLEVCGD-----VPQTEDEFET-----IEKFMD--SRI 202  
Qy 111 TKSEQGS-----DAEPKPIFVK-----QYLVKWKGLSVLHCSWVPEKEFKAY 155  
Db 203 GRKGATGASTIYAVEADGDPNAGFEKSKELGEIQYLIKWKGSWHHNTWETEETLK--- 259  
Qy 156 KSNHRLKTRVNNFHRQ-----MESFNSEDDDFVAIPEWTTVDRIACRE 200  
Db 260 QQNVKGMNKLNDYKKKQDQETKRWLNKASPEDVEYVNCQQLTDDLHKQYQIVERIIAHSN 319  
Qy 201 ED---GELEVLVYKELSYDECYWESESDIS-TFQNEIQRFKDVN-SRTRRSKDVD-HKR 254  
Db 320 QKSAAGYPDYCKWQGLPYSECSWEDGALIAKKFQARIDEYFYSRNSQKTTTFFKDCVKLKQ 379  
Qy 255 NPRDFQFHTPEFLKG-----LLHPYQLEGLNFLRFSWSKQTHVILADEMGLGKTIQISIA 310  
Db 380 RPR-FVALKKQPSYIGGHESLELRDYQLNGLNWLASHWCKNGSCILADEMGLGKTIQITIS 438  
Qy 311 LLASLFEENLI--PHLVIAPLSTLRNWEREFATWAPQNMVVMYFCGTAQARAVIREHEFY 368  
Db 439 FLNYLFHEHQLYGPFLLRVPLSTLTSWQREIQTWAPQMNNAVYLGDITSRNMRTHW-- 496  
Qy 369 SKDQKKIKKKSGQISSESKQRIKFDVLLTSYEMINLDSAVLKPKEWECMIVDEGHR 428  
Db 497 -----MHPQTKRLKFNILLTYEILLKDKSFLGLNWFVIGVDEAHLK 540

Qy 429 NKDSKLFSSLTQYSSNHRILLTGTPLONNLDELFMHFLDACKFGSLEEFQEEFKDINQ 488  
Db 541 NDDSLLYRTLIDFKSNHRLLLITGTPLONSLKLWLSLLHFIMPEKFSWEDFEEHKG-GR 599  
Qy 489 EEQISRLHKMLAPHLRRVKDVMKMPKPKKELILRVLDLSSLOKEYYKAIFTRNYQVLTK 548  
Db 600 EYGYASLHKELEPFLRRVKDVEKSLPAKVEQILRMEMSALQOYKYKWLTRNYKALS 659  
Qy 549 -KGAQISLNNIMMELRKVCCHPYMLEGVEP-VIHDANEAFKQLLESCKQLLDKMMVK 606  
Db 660 GSKGTSGLNIMMELKCCNHCYLKPPDDNEFYNKQALQHLIRSSGKLILLDKLLIR 719  
Qy 607 LKEQGHVLIYTQFQHMLDLEDYCTHKKQYERIDGKVGGAERQIRIDRFNAKNSNKF 666  
Db 720 LRERGNRVLIQSVRMVLDILAELVKYRQFPFQRLDGSIKGELRKQALDHFNAEGSEDFC 779  
Qy 667 FLLSTRAGGLGINLATADTVIYDSDWNPHADLOAMARAHRLGQTNKVMYRLINRG 726  
Db 780 FLLSTRAGGLGINLASADTVVIFDSDWNPQNDLOAQAARAHRIQKKQVNIYRLVTG 839  
Qy 727 ERMQLTKKKMVLEHLVVGKLKTQN-----INQELDDDIIRYSGKELFAS 771  
Db 840 EDILERAKKMMVLDHLVIQRMDDTGKTVLHTGSTPSSSTPPNKEBELSAILKFGAEELFKE 899  
Qy 772 EDDEAGKSGKIHDDAAIDKLDRDLVEAE--VSVDDEENGELKAFKAVANFEYIDEN 828  
Db 900 PEGEEQEP-----QEMDIDEILKRAETRENEPGPLTVGDE----LLSQFKVANFSNMDED 950  
Qy 829 EAAALEAQORVAAESKSSAGNSDRASYWEEL-----LKDKFELHQAELNALGKRKRS 880  
Db 951 D-IELEPER-----NSRNWEEIIPESQRRRIEERQKELEEIYMLPRMNC 996  
Qy 881 RKQVLSIEEDDLAGLEDVSSDGESEYAEESTDGEAAGQVQTGRRPYRRKGRDNLEPTPL 940  
Db 997 AKQ-----ISFNGSEGRRSRRYSGSDSITERKPKKRGPRPTIPRE- 1041  
Qy 941 MEGGRSFRVLGFNQSORAIFVQTLRMRYGAGNFDWKEFVPR---LKQKTPEEINEYGILF 997  
Db 1042 -----NIKGFSDAEIRRFITKSYKKFG-GPLERLDAVARDAELVDKSETDLRRLGELV 1092  
Qy 998 LKHIAEEIDENS-----PTFSDGVPKEGLRIEDVLVRIALLILVQEKVKF 1042  
Db 1093 HNGCIKALKDNSSGQERAGRLGKVKGPFTF-----RISGVQVNAKLVISHEELA- 1142  
Qy 1043 VEDHPGKPVFPSPRILER-----FPGLRSGKIWKEEHDKIMIRAVLKHGYGRWQAIVD 1094  
Db 1143 -----PLHKSIPSDPEERKRYVIPCHTKAAHFDIDWGKEDSNLLVIGIYEGYGSWEMIKM 1198  
Qy 1095 DKELGIQELI-----CKELNFPH---ISLSAAEQAGLQGGNGSGSNPGAQTNQN 1141  
Db 1199 DPDLSTQKILPDDPKKPOAKQLQTRADYLIKLLNKDLARKEAQRLAGAGNSKRRKTRN 1258  
RESULT 10  
US-10-437-963-201156  
; Sequence 201156, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966

```
; SEQ ID NO 201156
; LENGTH: 2192
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_96558C.1.pep
US-10-437-963-201156

Query Match      21.5%; Score 1548.5; DB 16; Length 2192;
Best Local Similarity 40.2%; Pred. No. 1.9e-106;
Matches 349; Conservative 163; Mismatches 239; Indels 117; Gaps 20;

QY 130 QYLVKWKGLSYLHCSWVPEKEFQKAYKSNHRLKTRVNNFHRQMESFNNSDDDFVAIRPEW 189
Db 491 EFLVKWVGKSNHNSWSESELKALAK-----RKLENY-----KAKY 527

QY 190 TTVDRIACREDEGELEYLVKYKELSYDECYWE--SESDISTFQNEIQRFKDVNSRT-RR 246
Db 528 GT-----DEIEEALIKWCALPYDECTWERLDEPTMVKYAHLVTQFKKFESQALDK 577

QY 247 SKVDVHKRNPRDFQF---DHTPEFLKGLLHPYQLEGLNFLRFSWSKQTHVILADEMGL 302
Db 578 DKGGS hak-PRHQEFENMLVEQPKELQGGMLFPHPQLEALNWLKWCYKSKNVILADEMGL 636

QY 303 GKTIOQSIALLASL---FEENLPHLVIAPISTLRNWEREFATWAPQMVVVMYFGTAQARA 359
Db 637 GKTVSACAFSLSLCCEYKINL-PCILVLPVLPSTMPNWMAEFASWAPHLNVVEYHGSARARS 695

QY 360 VIREHEFYLSKDQKIKKKKSGQISSSESQKRIKFDVLLTSYEMINLDSAVLKPIKWECM 419
Db 696 IIRQYEWHEGDASQMGKIKKSH-----KFNVLTTYYEMVLDAAYLRVSWEVL 744

QY 420 IVDEGHRLKNKDSKLPSSLTQYSSNHRILLTGTPQLQNNLDELFLMHFLDAGKFGSLEEF 479
Db 745 IVDEGHRLKNSSKLPSSLNTLSFQHRVLLTGTPQLQNNIGEMYNNLFLQPASFPSLASF 804

QY 480 QEEDKDIQNEEQISRLHKMLAPHLLRRVKDVMKMPKPKELIILRVDSLSSLOKYYKAIF 539
Db 805 EEKFNDLTTTEKVEELKNLVAPHMLRRLKKDAMQNI PPKTERMVPVELTSIQAEYYRAML 864

QY 540 TRNYQVLTK--KGGAQISLNNIMMELRKVCCHPYMLEGVEPVIHDANEAFKQLLESCGKL 597
Db 865 TKNYQVLRNIGKGAHQSLNIVMQLRKVCNHPYLPGPTEPGSGSPFELHMERIKASAKL 924

QY 598 QLLDKMMVKLKEQHRVLIYTOFHQMLDLLEDYCTHKKWQ-----YERIDGKVGGAERQI 652
Db 925 TLLHSMCLKILHKDHRVLIIFSOMTKLLDILEDYLT---WEFGPKTFERVDGSGVSAERQA 981

QY 653 RIDRFNAKNSKFCFLLSTRAGGLGINLATADTVIIYDSDNPHADLQAMARAHRLGQTN 712
Db 982 AIARFN-QDKSRFVFLSTRSCGLGINLATADTVIIYDSDFNPHADIQAMNRAHRIGQSN 1040

QY 713 KMIYRLINRGTIERRMQLTKKQWVLEHLVVGKLTQNIQNEELDDIIRYGSKELFAS- 771
Db 1041 RLLVYRLVVRASVEERILHLAKKMLDQLFVNKSE----SQKEVEDIIRWGTEELFRNS 1096

QY 772 -----EDDEAGKSGKIHYDDAAIDKLLDRD 796
Db 1097 DVAVKDNNEASGAKNDVAEVEFKHKRKTGGLGVYEDRCADGSAKFIWDENAITKLLDRS 1156

QY 797 LVEAEV-SVDDEEEENGFLKAFKVANF-EYIDENEAALAEQRV-----AAESKSSAG 847
Db 1157 NVPSTVAESTDGLDNDMLGTVKSIDWDELNDDPGATEDIPNIDNDGCGEQAASEAKQDAA 1216

QY 848 NSDRASYWEELLKDFELHQAEELNALGKRKRSPKQLVSI EEDDLAGLEDVSSDGDSEYE 907
Db 1217 NRVEENEWDKLLRVWEQYQTEEEASLGRGKRLKAVSYRET FATIPNEALS EDSDEE 1276

QY 908 AESTDGEAAGQGVQTGRPPYRRKGRDNL 935
Db 1277 PK-REYTAAGLALKEKYLRAQKERI 1303
```

```
RESULT 11
US-10-369-493-5100
; Sequence 5100, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5100
; LENGTH: 1465
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-5100

Query Match      21.4%; Score 1536.5; DB 15; Length 1465;
Best Local Similarity 30.0%; Pred. No. 8.1e-106;
Matches 454; Conservative 229; Mismatches 440; Indels 389; Gaps 56;

QY 21 DSDDDDF-----VPKKDRTF--EQVEAIVRTDAKENACQACGESTNLVSCNTCTYA 70
Db 169 DDIDDDVLEWDEEPAVPAEGPAILTETVERVIKW--RHGVPGATGSAT-----TC--- 217

QY 71 FHAKCLVPPLK DASVENWRCPECVSPNLNEIDKILDCEMRPTKSSSEQSSDAEPKPIFVKQ 130
Db 218 -----YNIADKGD PNDQIPGDKTEQ-----Q 238

QY 131 YLVKWKGLSYLHCSWVPEKEFQKAYKSNHRLKTRVNNF---HRQMESFNNSD-DFV--- 183
Db 239 FFKWTGWSHLHNTW--ESENSLALMNAKGLK-KVQNYVKQKEVEMWKR SADKEYIEFF 295

QY 184 -----AIRPEWTTVDRI LA-----CREEDGEL--EYLVKYKELSYDECYWESE---- 224
Db 296 ECEQQAEBELCEEYKKVERVVAHQTSRDRADAGSMATEYLIKWSGLPYSDCTWEDEKMA 355

QY 225 -SDISTFQNEIQRFKDVNSRTRRSKDVHDHKNRPDRDQFQDHTPEFLK-----GLLHPYQ 277
Db 356 PEQIKAYVHRIENLKSPN---KNSNVLRKR-PK-FEKFESMPDFLKTGDGESTHKL RDYQ 409

QY 278 LEGLNFLRFSWSKQTHVILADEMGLGKTIQSIALLASLFEENLI--PHLVIAPLSTLRNW 335
Db 410 LEGLNMVYAWCKGNSSILADEMGLGKTIQSIALLASLFRHYDLAGPYLVVVPVPLSTMAAW 469

QY 336 EREFATWAPQMVVVMYFGTAQARAVIREHEFEYLSKDQKIKKKKSGQISSSESQKRIKFD 395
Db 470 QKEFAQWAPFEMNLVVMGDVVS RDMIRQYEFV-----GGTKMKMIN 511

QY 396 VLLTSYEMINLDSAVLKPIKWECMIVDEGHRLKNKDSKLPSSLTQYSSNHRILLTGTP LQ 455
Db 512 AILTTYEILLKDKAFLSSIDWAALLVDEAHLRKNDESLLYKSLTQFRFNHKLITGTPLQ 571

QY 456 NNLDELFLMHFLDAGKFGSLEEFQEEFKDINQEEQISRLHKMLAPHLLRRVKKDVMMKOM 515
Db 572 NSLKELWALLHFIMPEKFCDCWEFEETAHNSN-HKGISALHKKLEPFLRRVKKDOVEKSL 630

QY 516 PPKKELILRVDSLSSLOKYYKAI FTRNYQVLTTKGGAQISLN---NIMMELRKVCCHPYM 572
Db 631 PPKTEQILRVDMTAHQQFYKWI LTKNYREL SK--GVKGSINGFVNLVWELKCCNHASL 688

QY 573 LEGVEPVIHDANEAFKQLLESCGKLQLLDKMMVKLKEQHRVLIYTOFHQMLDLLEDYCT 632
Db 689 TRYDHIYDDAQGRLLQOLLKSSGKLILLDKLCLRLKDKGHRVLIFSQMVVMDILQEY LQ 748

QY 633 HKKQYERIDGKVGGAERQIRIDRFNAKNSKFCFLLSTRAGGLGINLATADTVIIYDSD 692
```

Db 749 LRREPSQRLDGSWRADLRKQALDHYNAPGSTDFALLSTRAGGLINLATADTVIIIFDSD 808  
Qy 693 WNPHADLOAMARAHRLGQTNKVMYIRLINRGTIIEERMQLTKKQVLEHLVVGKLKTON- 751  
Db 809 WNPQNDLOAMSAHRIGQTKTVNIYRLVTKGSVEEIEIVERAKRKLVDHLVIOQMDTTGK 868  
Qy 752 -----INQEELDDIIRYGSKELFASEDDGAGSKGIHYDDAAIDKLDRDL 797  
Db 869 TVLSKNATASGVPFDKQELSAILKFGAVELFKEGEE-----QEPEVD--IDRIL 918  
Qy 798 VEAEVSVDBE--EENGFLKAFKVFANFEYIDENEAALAEQRAVAESKSSAGNSDRASYW 855  
Db 919 MGAETREAEVEVMKENELLSSFKYANFAIDEEKDIAAATDE-----W 960  
Qy 856 EELLKDK-----FELHQAEEELNALGKRKRKQ-LVSIIEEDDLAGLEDVSSDGDESYAE 909  
Db 961 AAIIPEEDNRILEERMKELAEWNLAPRQRKQPIQOVVEDD-----DGDDDEED 1011  
Qy 910 STDGEAAGQVQTRRPYRRKGRDNLEPTPLMEGEGRSFRVLGFNQSORAIFVQTLMRYG 969  
Db 1012 -----DTGKKK-KKKAUGNF-TIPEIKRPIKSFRR-----FSMPLNRLE 1048  
Qy 970 --AGNFWKEFVPRLKQKTFEINEYGIILFKHIAEIDENSPFSDGVPKEGLRIEDVL 1027  
Db 1049 EIAQDAELEHSTDEMKKLVESLSE-----ACKKADEFDSNEKNGDAGAAESEKKO---- 1100  
Qy 1028 VRIALLILVQEKVKF-----VEDHPGKVPFPSRILE-RFPGLRSGKI----- 1068  
Db 1101 -----IERKFKFHTCDVNLKQIERSHAELKPLHEILKSEETKTSFKPPANAKLQKGW 1152  
Qy 1069 ---WKEEHDKIMIRAVLKHGYGRWQAIIVDDKELGIGIELI----- 1104  
Db 1153 DVDWSRPDSSALLGVWKYGYGSWEAIKMDPTLGLADKIFIKDKTKKPGKNLQVRVDYL 1212  
Qy 1105 CKELNPFHISLSAAEQAGLQONGSGSNPGAQTNQNPGSVITGNNNASADGAQVNSMFY 1164  
Db 1213 LKLMSDKVKVTKTEKKERKRKADDPVPG-----PEKKRHTNNVPQEGEKKK--- 1259  
Qy 1165 YRDMQRRLVEFVKRVLLL--EKAMNYEYAEYYG--LGGSSSIP----- 1205  
Db 1260 -EKKEKNSSSLKDQLALLSIDKSL-----YGGALEDSSAKPFLCEVKLCMPVHKYM 1310  
Qy 1206 ---TEEPEA-----EPKIADTVGVSFIEVDDEMLDGLPKTDPTITSEEIMGAAVDNNQARV 1257  
Db 1311 KKLKEAQEAKNAQDAEAKYLTRLGDSFLENLETLIKPKPTN----- 1351  
Qy 1258 EIAQHYNQ----MCKLLDENARESVOAYVNNQPPSTKVNESFRALKSINGNINTILSITS 1313  
Db 1352 -IRKWNLYLWIFLCKFTLREPGEADRY-----RSITS 1383  
Qy 1314 DQSKSH-----EDDTKPDNLNNVEMKDTAETKPLRGVVDLNVVEGEENIAEASG-SV 1365  
Db 1384 DKHKNHHHHHHSKEEKPK----EAKDHKE-----RDREKDRERNRGERMDHGEGETSK 1434  
Qy 1366 DVKMEEAKEEEK 1377  
Db 1435 DHHREHKKDHHK 1446

RESULT 12  
US-09-849-602-16  
; Sequence 16, Application US/09849602  
; Publication No. US20030165834A1  
; GENERAL INFORMATION:  
; APPLICANT: Scanlan, Matthew J.  
; APPLICANT: Old, Lloyd J.  
; APPLICANT: Stockert, Elisabeth  
; APPLICANT: Chen, Yao-Tseng  
; TITLE OF INVENTION: Colon Cancer Antigen Panel  
; FILE REFERENCE: L0461/7105(JRV)  
; CURRENT APPLICATION NUMBER: US/09/849,602  
; CURRENT FILING DATE: 2001-05-04

; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 16  
; LENGTH: 1967  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-849-602-16  
  
Query Match 20.8%; Score 1492.5; DB 10; Length 1967;  
Best Local Similarity 30.4%; Pred. No. 2.6e-102;  
Matches 407; Conservative 228; Mismatches 423; Indels 279; Gaps 48;  
  
Qy 100 IDKILDCEMRPTKSSSEQSSDAEPKPIFVKQYLVKWKGLSYLHCSWVPEKEFQKAYKSNH 159  
Db 40 VEKIM--SSRSVKQKESGEEVE-----IEEFYVKYKNFSYLHCQWASIEDLEK----DK 88  
Qy 160 RLKTRVNNF---HRQMESFNNSDDFVAIRPEWTTVDRI---ACREEDGE--LEYLVKY 211  
Db 89 RIOQKIKRFKAKQGQNKFLSEIEDEL--FNPDYVEVDRIIMDFARSTDDRGEPTVTHYLVKW 146  
Qy 212 KELSVDCEYWESESDISTFQNEIQRFKDVNSRTRRSKDVHKNRPR-DFQQFDHTPEFL- 269  
Db 147 CSLPYEDSTWERRQDID--QAKIEEFKELMSREPETERVE--RPPADDDWKSESSEYKYN 202  
Qy 270 KGLLHPYQLEGLNFLRFSWSKQTHVILADEMGLKTIQSIALLASLFEENLI-PHLVIAP 328  
Db 203 NNKLREYQLEGVNWLLENWYNNRNCILADEMGLKTIQSIITFLYEIYLVKGIHGPFLVIAP 262  
Qy 329 LSTLRNWEREFATWAPQMVVYFGTAQARAVIREHEFVYLSKQKIKKKKSGQISSSEK 388  
Db 263 LSTIPNWEREFRTWT-ELNVVYVHGSGQASRRTIQLYEMYFKDPQGRVIKGS----- 312  
Qy 389 QKRIKFDVLLTSYEMINLDSAVLKPIKWECMIVDEGHRLLKNKDSKLSFSSLTOYSSNHRIL 448  
Db 313 ---YKFHAIITTFEMILTDCPELRNIPWRCVWIDEAHRLLKRNCKLLEGLKXMDLEHKVL 369  
Qy 449 LTGTPLQNNLDELFMLMHFLDACKFGSLEEFQEEFKDINQEEIISRLHKMLAPHLRRVK 508  
Db 370 LTGTPLQNTVEELFSLHLFLESFPSETTFMQEFGDLKTEEQVKLQAILKPMMLRLK 429  
Qy 509 KDVMKDMPPKKELILRVDLSSLOKEYYKAIIFTRNYQVLTKKGG-AQI-SLNNIMMELRKV 566  
Db 430 EDVEKNLAPKEETIIIEVELTNIQKYYRAILEKNFTFLSKGGQANVPNLLNTMMELRKC 489  
Qy 567 CCHPYMLEGVEPVI-----HDANEAFKQL---LESCGKLQLLDKQMVKLKEQGHVLI 616  
Db 490 CNHPYLINGAEKILEEFKETHNAESPDLQAMIOAAGKLVLDKLLPKLKAGGHRVLI 549  
Qy 617 YTOPQHMLDLEDYCTHKKWQYERIDGKVGGAERQIRIDRFNAKNSNKFCLLSTRAGGL 676  
Db 550 FSQWVRCLDILEDYLIQRRYPYERIDGRVGRNLRQAAIDRFSPKSDSDFVFLCLTRAGGL 609  
Qy 677 GINLATADTVIYDSDWNPHADLOAMARAHRLGQTNKVMYIRLINRGTIIEERMQLTKK 736  
Db 610 GINLTAADTCIIFDSDWNPQNDLQAQARCHRIGQSKSVKIYRLITRNSYEREMFDKASLK 669  
Qy 737 MVLEHLVVGKLLK-----TONINQBEELDDIIRYGSKELFASEDDGAGSKGIHYDDAA 788  
Db 670 LGLDKAVLQSMGSGRENATNGVQOLSKEIEIDLRLKGAYGALMDEDEGSK-----FCEED 724  
Qy 789 IDKLLDRDLVEAEVSVDDDEENGFLKAFKVFANFEYIDENEAALAEQRAVAESKSSAGN 848  
Db 725 IDQIL---LRRTHITITIESE---GKGSTFAKASF-----VASGNRTDI-S 762  
Qy 849 SDRASYWEEELLKDKFELHQAEEELNALGKRK-----RSRKO---LVSIIEEDDLAGLEDV 898  
Db 763 LDDPNFWQKWAK-KAEL----DIDALNGRNNLVIDTPRVKQTRLYSAVKEDELMFSD 817  
Qy 899 SSDGDESYEAESTDGEAAGQVQTRRPYRRKGRDNLEPTPLMEGEGRSFRVLGFNQSQ- 957  
Db 818 ESDSEKPCAKPRRPQDKSQG-----YARSECFRVEKNLLVYGWGRWTDILSHGRYKR 870  
Qy 958 -----RAIFVQTLNRY-----GAGNFDWKEFVPRLKQKTFEINEYGI----- 995



Db	871	QLTEQDVETICTILVYCLNHYKGDENIKSFIWDLITPTADGQTRALVNHSGLSAPVPRG	930
QY	996	-----LFLKHIAEID-----ENSPFSDGVPKEGLR--IEDVLVRIALLILVQEKV	1040
Db	931	RKGKKVKAQSTQPVVQADWLASCNPDALFQEDSYKKHLKHHCNKVLLRVRLMYLQREV	990
QY	1041	-----KFVE--DHPGKPVFPPSRILERFPGLRSGKIWKEEHDKIMTRAVLKHGYGRWQAI	1092
Db	991	IGDQADKILEGADSEADVW---IPEPFHAEVPADWWDKEADKSLIGVFKHGYEKYNSM	1047
QY	1093	VDDKELGIELICKELNFPFHISLSAAEQAG--LQGQNGSGG-----	1131
Db	1048	RADPALCFLE---RVGMPDAKAIAAEQRGTDMLADGGDGGEFDRDEBDPEYKPTRTPFK	1103
QY	1132	-----SNPGAQTNQNPGSVITGNNNASADGAQVNSMFY-----	1164
Db	1104	DEIDEFANSPEDEKESMEIHATKHSSES--NAELGQLYWPNTSTLTTRLRLITAYQRS	1161
QY	1165	YRDMQRRLLVEFK-----KRVLLLEKAMNYEYABE--YYGLGGSSSI	1204
Db	1162	YKQOMRQEALMKTDRRRRRPREEVRALEAREAIISEKRQKWTREEDAFY-----	1213
QY	1205	PTEEPEAEPKIADTVGVSFIEVDDEMGLDKPTDPTSEEIMGAADVNNQARVEIAQHYN	1264
Db	1214	-----RVVSTFGVIF-----DPVKQQ-----FDWNQFRA-----	1237
QY	1265	QMCKLLDENARESVOAY	1281
Db	1238	--FARLDKKSDESLEKY	1252
RESULT 13			
US-10-369-493-2310			
; Sequence 2310, Application US/10369493			
; Publication No. US20030233675A1			
; GENERAL INFORMATION:			
; APPLICANT: Cao, Yongwei			
; APPLICANT: Hinkle, Gregory J.			
; APPLICANT: Slater, Steven C.			
; APPLICANT: Goldman, Barry S.			
; APPLICANT: Chen, Xianfeng			
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF			
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES			
; FILE REFERENCE: 38-10(52052)B			
; CURRENT APPLICATION NUMBER: US/10/369,493			
; CURRENT FILING DATE: 2003-02-28			
; PRIOR APPLICATION NUMBER: US 60/360,039			
; PRIOR FILING DATE: 2002-02-21			
; NUMBER OF SEQ ID NOS: 47374			
; SEQ ID NO 2310			
; LENGTH: 1407			
; TYPE: PRT			
; ORGANISM: Schizosaccharomyces pombe			
; FEATURE:			
; NAME/KEY: unsure			
; LOCATION: (1)..(1407)			
; OTHER INFORMATION: unsure at all Xaa locations			
US-10-369-493-2310			
Query Match			
Best Local Similarity 20.5%; Score 1473; DB 15; Length 1407;			
Matches 429; Conservative 215; Mismatches 437; Indels 382; Gaps 42;			
QY	12	SDRKPVYNLDD-----SDDDDF--VPKKDRTFEQVEAIVRTDAKENACQACGES	58
Db	118	SRHRTIRDLDDAEASTSEESDSSYGGTPKKRSRQKSNTRYVQDEIRFSSRNKGVN	177
QY	59	TNLVSCNTCTYAFHAKCLVPPLKDAVENW-----RCPECVSPLNE-----IDKILD	105
Db	178	YN-----EDAYFESFEEEEEMYEYATEVSEEPEDTRAIDVVLD	217
QY	106	CEMRPTKSEQSSDAEPKPIFVKQYLVKWKGLSYLHCSWVPEKEFQKAYKSNHRLKTRV	165

Db	218	HRLIEGHGSGTSPSEDYE-----FLIKWVNFSLHCTWEPYNNI-----SMIRGSKV	264
QY	166	NNFHRQM-----ESFNNSDEDFVAI-----RPEWTTVDRILACR--EEDGELE	206
Db	265	DNHIKQVILLDREIREDPPTTTTREDIEAMDIEKERKRENYEYKQVDRIIVAKHLNSDGSVE	324
QY	207	YLVVKYKELSYDECYWESESDISTF-QNEIQRFKD-VNSRTRRSKDVVDHKNRPRDFQQFDH	264
Db	325	YLVKWKQLLYDFCTWEASSIIIEPIAAATEIOAQFEREESALSPRGTYNGNSRPKYRKLEQ	384
QY	265	TPEFLK-GLLHPYQLEGLNFLRFSWSKQTHVILADEMGLKTTIQSIALLA-----SLFEEN	319
Db	385	QPSYITGGELRDFQLTGVNMWMAYLWHKKNENGILADEMGLGKTIVQTVAFLSYLAHSLRQHG	444
QY	320	LIPHLVIAPLSTLRNWEREFATWAPQMNVMVYFGTAQARAVIREHEFYLSKDQKIKKKK	379
Db	445	--PFLVVVPLSTVPAWQETTLALWASDMNCISYLGNTTSRQVIRDYEFYVDGTQK-----	496
QY	380	SGQISSESQKRIKFDVLLTTSYEMINLDSAVLKPIKWECEMIVDEGHRLLKNKDSKLFSSLT	439
Db	497	-----IKFNLLTTTYEYVLKDRSVLSNIKWQYMAIDEAHRLKNSESSLYEALS	544
QY	440	QYSSNHRILLTGTPLONNDELFMLMHFLDAGKFGSLEEFQEEFKDINQEEQISRLHKML	499
Db	545	QFKNSNRLLITGTPLONNIRELAALVDFLMPGKFEIREEINLEAPDEEQEAYIRSLOEHL	604
QY	500	APHLRRVKDVMKDMPPKKELILRVDLSSLOXEYKAIFFRNYQVLTG--KGAQOISLN	557
Db	605	QPYILRRLLKDKVEKSLPSKSERILRVELSDLQMYWYKNILTRNYRVLTSQISSGSQISLL	664
QY	558	NIMMELRKVCCHPYMLEGVE----PVIHD---ANEAFKQLLESCKQLQLLDKMMVKLKEQ	610
Db	665	NIVVELKKASNHPYLFDFGVEESWMQKINSQRRRDEVLKLIMNSGKMVLLDKLLSRLRRD	724
QY	611	GHRVLIYTFQHMLDLLEDYCTHKWQYERIDGKVGGAERQIRIDRFNAKNSNKFCLLS	670
Db	725	GHRVLIFSQVMRMLDILGDYLSLRGYPHQRLDGTVPAAVRRTSIDHFNAPNSPDFVFLLS	784
QY	671	TRAGGLGINLATADTVIYDSDMNPHADLQAMARAHRLGQTNKVMYRNLINRGTIERMM	730
Db	785	TRAGGLGINLMTADTVIIFDSDMNPQADLQAMARAHRIQGNHVMVYRLLSKDTIEEDVL	844
QY	731	QLTKKKMVLHLVVG-----KLKTQINQBEELDDIIRYGSKELFASEDDEAGSGKI	782
Db	845	ERARRKMILEYAIISLGVTDKQKNSKNDKFSAEELSAILKFGASNMFAENNQ-----	897
QY	783	HYDDAAIDKLLDRDLVEAEVSVDDDEENG-----FLKAFKVANFE-----	823
Db	898	-----KKLEDNMNLDLEIHAEDHDTSDNDVGGASMGGEFLKQFEVTDYKADVSWDDII	950
QY	824	-----YIDENEAALAEQ-----RVAAESKSSAGNSDRASYWEELL	859
Db	951	PLTEREKFEEDRLREEEALKQEIELSSRRGNRPYPSSAVESPYSGTSEKSK-KQML	1009
QY	860	KDKFELHQAELNALGKRKRSQLVSIIEE--DDLAGEDEVSSDGDSEYAEASTDGEAAG	917
Db	1010	KDEVILLE--KEIRLL---YRAMIRYGSLEHRYNDIVKYADLTITQDAHVIKKIAADLVAS	1064
QY	918	QGVQTRRRPYRRKGRDNLEPTPLMEGEGRSFRVLGFNQSORAIFVQITLMRYGAGNFDWKE	977
Db	1065	RKAVSAAEKDLSNDQSN-----NKSSRKALLITEK--GVKNINAET	1103
QY	978	FVPRLLKQKTFEINEYGILFLKHIAEEIDENSPTFSDGVPKPEGLRIEDVLVRIALLLVQ	1037
Db	1104	LVQRL-----NDLDILY-----DAMPTSGYSNFQIPMHV-----	1132
QY	1038	EKVKFVEDHPGKPVFPFSRILERFPGLRS-----GKIWKEEHDKIMIRAVLKHGYGRWQAI	1093
Db	1133	-----RSVHGWSCQWGPREDSMLLSGICKHGFCAWLEIR	1166
QY	1094	DDKELGIELICKELNFPHISLSAAEQAGLQGNQSGSGSNPGAQTNQNPGSVITGNNNAS	1153

Db 1167 DPELKMCKIFLE-----DTKQTDNSVPKOKENKE 1197

Qy 1154 ADGAQVNSMPYRDMQRRLLVEFVKRVLLEKAMNYEAYEYGLGSSSIPTTEEPAEP 1213

Db 1198 ---KKVPS-----AVHLVRRGEYLLSALREH---HQFNGKSSPAISTN----- 1235

Qy 1214 KIADTVGSFIEVDMDGLPKTDPITSEEIMGAADVNNQARVEIAQHYNQMCKLLDEN 1273

Db 1236 -----GKTQP-----KKQTAN 1246

Qy 1274 ARESVQAYVNN-----QPPSTKVNESFRALKSINGNINT-----ILSITSDQKSHE 1320

Db 1247 RRQSGKPNVSAQKIESATRTSPSPAISESRKKPSSKDTKIETPSREQRSQTASPVKSEK 1306

Qy 1321 DDTKPDNNVEMKDTAEETKPLR 1343

Db 1307 DQGNVSLNABQKARCKELMYPVR 1329

RESULT 14

US-08-973-363-21

; Sequence 21, Application US/08973363

; Publication No. US20030191297A1

; GENERAL INFORMATION:

; APPLICANT:

; TITLE OF INVENTION: AVIAN GHD GENES AND THEIR USE IN METHODS FOR

; TITLE OF INVENTION: SEX IDENTIFICATION IN BIRDS

; NUMBER OF SEQUENCES: 39

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: WENDEROTH, LIND & PONACK, L.L.P.

; STREET: 2033 K. Street, N.W., Suite 800,

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20006

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/973.363

; FILING DATE:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: WO PCT/GB96/01341

; FILING DATE: 05-JUN-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB 9511439.3

; FILING DATE: 06-JUN-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Warren M Cheek, Jr.

; REGISTRATION NUMBER: 33,367

; REFERENCE/DOCKET NUMBER: 263/PPNTR1172US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202)-721-8200

; TELEFAX: (202)-721-8250

; TELEX:

; INFORMATION FOR SEQ ID NO: 21:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1467 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; FEATURE:

; NAME/KEY: Peptide

; LOCATION: 1..1467

; OTHER INFORMATION: /note= "The sequence beginning at 1

; OTHER INFORMATION: corresponds to 2654 and that ending at 1467 corresponds to

; OTHER INFORMATION: 4120"

US-08-973-363-21

Query Match 20.3%; Score 1461.5; DB 8; Length 1467;

Best Local Similarity 32.1%; Pred. No. 3.6e-100;

Matches 410; Conservative 217; Mismatches 403; Indels 249; Gaps 46;

Qy 7 RLRIIRSDRKPVYNLDDSDDDDFVPKKORTFEQVEAIVRTDAKENACQACGESTNLVSCNT 66

Db 126 RFSNRQKNTVYNIDYSDDDLLESEDD--YGSEAL---SEENVHEA-----SANP 171

Qy 67 CTYAFHAKCLV--PPLKDASVENWRCPECVSPNLNEIDKILDCBMRPTKSSSEQSSDAEPK 124

Db 172 QPEDFHGIDIVINHRLKTSLEBKVLEKTPVDLN-----NCKEN----- 210

Qy 125 PIFVKQYLVKWKGLSYLHCSWVPEKEFKAYKSNHRLK--TRVNNFHRQ-----ME 173

Db 211 ----YEFLIKWTDESHLNTW-----ETYESIGQVRLKRLDNYCKQFIEDQQVRLD 259

Qy 174 SFNNSD-----DFVAIR-----PEWTTVDRIACRE---EDG--ELEYLVKYKELSYDEC 219

Db 260 PYVTAEDIEIMDMERERRLDEFEFHVPERIIDSORASLEDGTSQLYLVKWRRLNYDEA 319

Qy 220 YWESESDISTFQNE-IQRFKD-VNSRTRRSKVDVHKRNPRDFQQFDHTPEFLK-GLLHPY 276

Db 320 TWENATDIVKLAPEQVKHFQNRKNSKILPOYSSNYTSQRPRFEKLSVQPPFIKGGELRDF 379

Qy 277 QLEGLNFLRFSWSKQTHVILADEMGLGKTIQSIALLASLF--EENLPHLVIAPLSTLRN 334

Db 380 QLTGINWMAFLWSKGDNGILADEMGLGKTVQTVAFISWLIPARRQNGPHIIVVPLSTMPA 439

Qy 335 WEREFPATWAPQMNVMYFGTAQARAVIREHEFYLSKDQKIKKKKSGQISSESKQRIKF 394

Db 440 WLDTFEKWAPDLNCICYMGNOKSRDTIREYEFYTNPRAK-----GKKTMKF 485

Qy 395 DVLLTSYEMINLDSAVLKPIKECMIVDEGHRLKNKDSKLFPSSLTQYSSNHRILLTGTP 454

Db 486 NVLLTTYEYILKDRAELGSIKQFMVDEAHRLKNAESSLSNFKVANRMLITGTP 545

Qy 455 QNNLDELFLMHFLDAGKFSLEEFQEEFKDINQEEQISRLHKMLAPHLRLRRVKQVMKD 514

Db 546 QNNIKELAAALVNFMPGRFTIDQEIQDFENQDEEYIIDLHRRIOQPFILRRLKKDVEKS 605

Qy 515 MPPKKELILRVDLSSSQKEYYKAIFTRNYQVLT--KKGGAQISLNNIMMELRKVCCHPYM 572

Db 606 LPSKTERILRVELSDVQTEYYKNILTKNYSALTAGAKG-HFSLNIMNELKASNHPLY 664

Qy 573 LEGVEPVI-----HDANAEAFKQLLESCGKLQLDKMVKLKEQGHRLVIYTFQFQHML 624

Db 665 FDNAEERVQKFGDGKMTRENVLRLGLIMSSGKMWLLDQLTRLKKDGHRLVIFSQVMRL 724

Qy 625 DLLEDYCTHKQOYERIDGKVGGAERQIRIDRENAKNSKFCFLLSSTRAGGLINLATAD 684

Db 725 DILGDYLSIKGINFORLDGTGPSAQRRIISIDHFNPSDNDPFLSTRAGGLINLMTAD 784

Qy 685 TVIYSDWNPHADLOAMAPAHRLGQTNKVMYRLINRGTEERMQLTKKQVLEHLVV 744

Db 785 TVVIFSDWNPQADLOAMARAHRIQGNHVMYVRLVSKDITVEEVLERARKOMILEYAI 844

Qy 745 G-----KLKTQININQEEIIRYGSKELFASEDDACKSGKIHYYDDA---AIDKLL 793

Db 845 SLGVTGNGKYTKKNEPNAGELSAILKFGAGNMFTATDNQK-KLEDNLDDVLNHAEDHVT 903

Qy 794 DRDLVEAEVSVDDDEENGFLKAFKVFANF-----EYIDENEAALAEQRAVAESKSSA 846

Db 904 TPDLGESH---LGSEE---FLKQFEVTDYKADIDWDIIPPEELKKLQ----- 945

Qy 847 GNSDRASYWEELLKDKFELHQAEEELNALGKRKRSRKQLV-SIEEDDLAGLEDVSSDGDES 905

Db 946 -----DEEQKRKDEEYVKEQLEMMNRRDNALKKKNVSGDGTGTA---ANSDSDD- 991

Qy 906 YEAESTDGEAAGQGVQGTGRPPYRRKGRDNLEPTPLMEGGRSF--RVLGFNQ----- 955

Db 992 ---DSTS-----RSSRRRARAN--DMDSIGSESEVRALYKAILKFGNLKELDEL 1035

Qy 956 -SQRAIFVQTLRMRYGAGNFDWKEFVPRLLKQKTFEINNEYGILFLKHAIEEIDENSPTF-- 1012

Db 1036 IADGTLPVKSFKEYGE---TYDEMMEAADCVHEEKN-----RKEILEKLEKHATAYRA 1087

Qy 1013 ---SDGVPKEGLRIEDVLVRIALLILVQEKVKFVEDHPG-KPVFPSPRILERFPGLRSGK- 1067

Db 1088 KLSGEIKAENQPKDNPLTRLSL--KKREKKAFLFNFKGVKSLNAESLLSRVEDLKYLKN 1145

Qy 1068 -----IWKEEHDKIMIRAVLKHGYGRWQAIVDDKELG 1099

Db 1146 LINSNYKDDPLKFSLGNNTPKPVQNWSSNWTKEEDEKLLIGVFKYGYGSWTQIRDDPFLG 1205

Qy 1100 I-QELICKELNFP--HISLSAAEQAGLQCGQNGSG--GSNPGAQTNQNPQSV----- 1145

Db 1206 ITDKIFLNEVHNPNVAKKSASSSDTTPTPSKKGKGITGS-----SKKVPGAHLGRRVDYL 1260

Qy 1146 ---ITGNNNASADGAQVNS 1161

Db 1261 LSFLRGGLNTKSPSADIGS 1279

RESULT 15

US-10-369-493-1775

; Sequence 1775, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 1775

; LENGTH: 1468

; TYPE: PRT

; ORGANISM: Saccharomyces cerevisiae

US-10-369-493-1775

Query Match 20.3%; Score 1461.5; DB 15; Length 1468;

Best Local Similarity 32.1%; Pred. No. 3.6e-100;

Matches 410; Conservative 217; Mismatches 403; Indels 249; Gaps 46;

Qy 7 RLRIRSDRKPVYNLDDSDDDFVPKKDRTPQEVEAIVRTDAKENACACGESTNLVSCNT 66

Db 126 RFSNRQNTVNYNIDYSDDDLSEDD--YGSEAL----SEENVHEA-----SANP 171

Qy 67 CTYAFHAKCLV--PPLKDASVENWRCPCEVSPLEIDKILDCEMRPTKSSEQGSSDAEPK 124

Db 172 QPEDFHGIDIVINHRLKTSLEEGKVLEKTVPDLN-----NCKEN----- 210

Qy 125 PIFVKQYLVKWKGLSYLHCSWVPEKEFQKAYKSNHRLK--TRVNNFHRQ-----ME 173

Db 211 ----YEFLIKWTDESHLNTW-----ETYESIGQVRGLKRLDNYCKQFIIEDQQVRLD 259

Qy 174 SFNNSED----DFVAIR-----PEWTTVDRIACRE---EDG--ELEYLVKYKELSYDEC 219

Db 260 PYVTAEDIEIMDERRRLDEFEEFHVPERIIDSQRASLEDGTSQLQYLVKWRRRLNYDEA 319

Qy 220 YWESESDISTFQNE-IQRFKD-VNSRTRRSKDVDRHKNRPDRFQOQDHTPEFLK-GLLHPY 276

Db 320 TWENATDIVKLAPEQVKHFQNRNSKILPQYSSNYTSQRPREFKLSVQPPFIKGGEELRDF 379

Qy 277 QLEGLNFRFSWSKQTHVILADEMGLGKTIQSIALLASLF--EENLPHLVIAPLSTLRN 334

Db 380 QLTGINWMAFLWSKGDNGILADEMGLGKTVQTVAFISWLIFARRQNGPHIIVVPLSTMPA 439

Qy 335 WEREFATWAPQNVVMYFGTAQARAVIREHEFYLSKDQKIKKKKSGQISSSESQKRIKF 394

Db 440 WLDTFEKWAPDLNCICYMGNQSRDRTIREYEFYTNPRAK-----GKKTMKF 485

Qy 395 DVLTSTYEMINLDSAVLKPIKWECMIVDEGHRLLKNKDSKLFSSLTQYSSNHRILLTGTP 454

Db 486 NVLLTTYEYILKORAEKSGIKWQFMAVDEAHRLLKNAESSLSNFKVANRMLITGTP 545

Qy 455 QNNLDELFLMHFLDAGKFGSLEEFQEEFKDINQEEQISRLHKMLAPHLRLRRVKQDVMD 514

Db 546 QNNIKELAALVNFMPGRFTIDQIBIDFENQDEEQEEYIHDLHRRIQPFILRRLKQDVEK 605

Qy 515 MPPKKELILRVDLSSLQKEYKKAIFTRNYQVLT--KKGGAQISLNNIMMELRKVCCHPYM 572

Db 606 LPSKTERILRVELSDVQTEYYKNILTKNYSALTAGAKGG-HFSLNIMNMLKASNHPYL 664

Qy 573 LEGVEPVI-----HDANEAFKQLLES CGKLQLLDKMVKLKEQGHRVLIYTQFQHML 624

Db 665 FDNABERVLOKFGDGKMTRENVLRGLIMSSGKMVLDDQLLRLKDKDGHVLIQFQVRL 724

Qy 625 DLLEDYCTHKKWQYERIDGKVGGAERQIRIDRFNAKNSKFCFLSTTRAGGLGINLATAD 684

Db 725 DILGDYLSIKGINFQRLDGTVPQAORRISIDHFNSPDSNDFVLLSTRAGGLGINLMTAD 784

Qy 685 TVIYDSDNPHADLQAMARHRLGQTNKVMYRLINRGTIETERMMLTKKMVLEHLVV 744

Db 785 TVVIFDSDNWPNQADLQAMARAHRIQKNHVMYRLVSKDTVEEVLERARKKMILEYAI 844

Qy 745 G-----KLKTQINQEEELDDIIRYSGKELFASEDDDEAGSGKIHYYDDA---AIDKLL 793

Db 845 SLGVTGDNKYTKKNEPNAGELSAILKFGAGNMFTATDNQK-KLEDNLDDVLNHAEDHVT 903

Qy 794 DRDLVEAEVSVDDDEENGFLKAPKVFANF-----EYIDENEAALAEQVAAESKSSA 846

Db 904 TPDLGESH---LGEE---FLKQFEVTDYKADIDWDDIIPEELKKLQ----- 945

Qy 847 GNSDRASYWEELLKDKFELHQAELSINALGKRKRQV--SIEEDDLAGLEDVSSDGDDES 905

Db 946 -----DEEQKRKDEEYVKEQLEMMNRDRNALKKIKNSVNGDGT-----ANSDD- 991

Qy 906 YEAESTDGEAAGQGVQTGRRPYRRKGRDNLEPTPLMEGGRSF--RVLGFNQ----- 955

Db 992 ---DSTS-----RSSRRARAN--DMDSIGESEVRALYKAILKFGNLKEILDEL 1035

Qy 956 -SQRAIFVQTLMYGAGNFDWKEFVPRLKQKTFEINEYGILFLKXIAEEIDENSPTF-- 1012

Db 1036 IADGTLPVKSFKEYGE---TYDEMMEAADCVHEEKN-----RKEILEKLEKHATAYRA 1087

Qy 1013 ---SDGVPKEGLRIEDVLVRIALLILVQEKVKFVEDHPG-KPVFPSPRILERFPGLRSGK- 1067

Db 1088 KLSGEIKAENQPKDNPLTRLSL--KKREKKAFLFNFKGVKSLNAESLLSRVEDLKYLKN 1145

Qy 1068 -----IWKEEHDKIMIRAVLKHGYGRWQAIVDDKELG 1099

Db 1146 LINSNYKDDPLKFSLGNNTPKPVQNWSSNWTKEEDEKLLIGVFKYGYGSWTQIRDDPFLG 1205

Qy 1100 I-QELICKELNFP--HISLSAAEQAGLQCGQNGSG--GSNPGAQTNQNPQSV----- 1145

Db 1206 ITDKIFLNEVHNPNVAKKSASSSDTTPTPSKKGKGITGS-----SKKVPGAHLGRRVDYL 1260

Qy 1146 ---ITGNNNASADGAQVNS 1161

Db 1261 LSFLRGGLNTKSPSADIGS 1279

Search completed: September 14, 2004, 02:12:18

Job time : 98 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 14, 2004, 01:59:21 ; Search time 53 Seconds  
(without alignments)  
2511.869 Million cell updates/sec

Title: US-10-049-137-2  
Perfect score: 7187  
Sequence: 1 MSSSLVERLRIRSDRKPVYNL.....VDVKMEEAKEEEKPKMVMVD 1384

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query %		DB ID	Description
	Score	Match Length		
1	7187	100.0	1384 2 T52301	GYMNOS/PICKLE prot
2	6921.5	96.3	1359 2 B84645	hypothetical prote
3	3114	43.3	1067 2 T06312	hypothetical prote
4	2048.5	28.5	1829 2 T34239	hypothetical prote
5	2032	28.3	1787 2 T20160	hypothetical prote
6	1934.5	26.9	1388 2 T17269	hypothetical prote
7	1572.5	21.9	1711 1 A47392	chromodomain-helic
8	1565.5	21.8	1738 2 C84507	hypothetical prote
9	1536.5	21.4	1465 2 T23056	chromodomain helic
10	1520.5	21.2	1883 2 T13944	chromodomain helic
11	1473	20.5	1388 2 T38720	chromodomain helic
12	1461.5	20.3	1468 1 S30818	hypothetical prote
13	1402	19.5	1373 2 T50107	chromodomain helic
14	1360	18.9	1367 2 T43334	chromodomain helic
15	1332.5	18.5	2957 2 T33152	hypothetical prote
16	1139	15.8	976 2 S35457	SNF2 protein homol
17	1139	15.8	1120 2 S67208	hypothetical prote
18	1135.5	15.8	1143 2 S46122	SNF2 protein homol
19	1131	15.7	1027 2 A56533	chromatin remodell
20	1126	15.7	971 2 S44645	hypothetical prote
21	1120	15.6	1359 2 S49883	nuclear protein ST
22	1114	15.5	1422 2 T18404	chromatin remodell
23	1101.5	15.3	769 2 S35458	SNF2 protein homol
24	1095.5	15.2	1703 2 S15047	SNF2 protein - yea
25	1093.5	15.2	1474 2 T20488	hypothetical prote
26	1052	14.6	1199 2 T37561	probable transcrip
27	1038.5	14.4	1638 2 A42091	transcription acti
28	1030.5	14.3	1022 2 I53078	homeotic gene regu
29	1027	14.3	1647 2 S45252	SNF2beta protein -

ALIGNMENTS

RESULT 1

T52301  
GYMNOS/PICKLE protein [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 20-Apr-2001  
C;Accession: T52301  
R;Eshed, Y.; Baum, S.F.; Bowman, J.L.  
Cell 99, 199-209, 1999  
A;Title: Distinct mechanisms promote polarity establishment in carpels of Arabidopsis thaliana  
A;Reference number: Z26025; MUID:20004396; PMID:10535738  
A;Accession: T52301  
A;Status: preliminary; translated from GB/EMBL/DBDJ  
A;Molecule type: mRNA  
A;Residues: 1-1384 <ESH>  
A;Cross-references: EMBL:AF185578; PIDN:AAF07084.1  
A;Experimental source: cultivar Colombia

Query Match 100.0%; Score 7187; DB 2; Length 1384;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1384; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MSSSLVERLRIRSDRKPVYNLDDDDDFVPKKDRTFEQVEAIVRTDAKENACQACGESTN	60
Db	1	MSSSLVERLRIRSDRKPVYNLDDDDDFVPKKDRTFEQVEAIVRTDAKENACQACGESTN	60
QY	61	LVSCNTCTYAFHAKCLVPPLKDASVENWRCPECVSPLNIDKILDCEMRPTKSSEQSSD	120
Db	61	LVSCNTCTYAFHAKCLVPPLKDASVENWRCPECVSPLNIDKILDCEMRPTKSSEQSSD	120
QY	121	AEPKPIFVKQYLVKWKGLSYLHCSWVPEKEFQKAYKSNHRLKTRVNNFHRQMESFNNSD	180
Db	121	AEPKPIFVKQYLVKWKGLSYLHCSWVPEKEFQKAYKSNHRLKTRVNNFHRQMESFNNSD	180
QY	181	DFVAIRPEWTTVDRILACREEDGELEVLVKYKELSYDECYWESESDISTFQNEIQRFKDV	240
Db	181	DFVAIRPEWTTVDRILACREEDGELEVLVKYKELSYDECYWESESDISTFQNEIQRFKDV	240
QY	241	NSRTRRSKDVHDHKNRPRDFQFQDHTPEFLKGLLHPYQLEGNLFRLFWSKQTHVILADEM	300
Db	241	NSRTRRSKDVHDHKNRPRDFQFQDHTPEFLKGLLHPYQLEGNLFRLFWSKQTHVILADEM	300
QY	301	GLGKTIQSIALLASLFEENLIPLHVIAPLSTLRNWEREFATWAPQMVVMYFGTAQARAV	360
Db	301	GLGKTIQSIALLASLFEENLIPLHVIAPLSTLRNWEREFATWAPQMVVMYFGTAQARAV	360
QY	361	IREHEFYLSKDQKKIKKKKSGQISSESKQKRIKFDVLLTSYEMINLDSAVLKPIKWECEMI	420
Db	361	IREHEFYLSKDQKKIKKKKSGQISSESKQKRIKFDVLLTSYEMINLDSAVLKPIKWECEMI	420
QY	421	VDEGHRLLKNKDSKLFSSLLTQYSSNHRILLTGTPLQNNLDELFMFLDAGKFGSLEEFQ	480
Db	421	VDEGHRLLKNKDSKLFSSLLTQYSSNHRILLTGTPLQNNLDELFMFLDAGKFGSLEEFQ	480

481 EEFKDINQEEQISRLHKMLAPHLLRRVKKDVNKMPPPKKELILRVLSSLOKEYYKAIFT 540  
|||||  
481 EEFKDINQEEQISRLHKMLAPHLLRRVKKDVNKMPPPKKELILRVLSSLOKEYYKAIFT 540  
|||||  
541 RNYQVLTKKGAQISLNNIMMELRKVCCHPYMLEGVEPVIIHDANEAFKQLL 600  
|||||  
541 RNYQVLTKKGAQISLNNIMMELRKVCCHPYMLEGVEPVIIHDANEAFKQLL 600  
|||||  
601 DKMVKLKEQGHRLVITYTQFQHMLDLEDYCTHKKWQYERIDGKVGGAERQIRIDRFNAK 660  
|||||  
601 DKMVKLKEQGHRLVITYTQFQHMLDLEDYCTHKKWQYERIDGKVGGAERQIRIDRFNAK 660  
|||||  
661 NSNKFCLLSTRAGGLGINLATADTVIIYDSWNPHADLQAMARAHRLGOTNKVMYIRLI 720  
|||||  
661 NSNKFCLLSTRAGGLGINLATADTVIIYDSWNPHADLQAMARAHRLGOTNKVMYIRLI 720  
|||||  
721 NRGTIERMMQLTKKKMWLEHLVVGKLTQNTINQEBELDDIIRYGSKELFASEDDGAKSG 780  
|||||  
721 NRGTIERMMQLTKKKMWLEHLVVGKLTQNTINQEBELDDIIRYGSKELFASEDDGAKSG 780  
|||||  
781 KIHYYDDAAIDKLLDRDLVEAEVSVDDDEENGFLKAFKVANFEYIDENEAALAEQRAA 840  
|||||  
781 KIHYYDDAAIDKLLDRDLVEAEVSVDDDEENGFLKAFKVANFEYIDENEAALAEQRAA 840  
|||||  
841 ESKSSAGNSDRASYWBEELLKDKFELHQAEEELNALGKRKRSRKQLVSIIEEDLAGLEDVSS 900  
|||||  
841 ESKSSAGNSDRASYWBEELLKDKFELHQAEEELNALGKRKRSRKQLVSIIEEDLAGLEDVSS 900  
|||||  
901 DGDESYEAESTDGEAAGQGVQTGRRPYRRKGRDNLEPTPLMEGGRSFRVLGPNQSORAI 960  
|||||  
901 DGDESYEAESTDGEAAGQGVQTGRRPYRRKGRDNLEPTPLMEGGRSFRVLGPNQSORAI 960  
|||||  
961 FVQTLMRYGAGNFDWKEFVPRLKQKTFEEINEYGITFLKHIAEEIDENSPTFSDGVPKEG 1020  
|||||  
961 FVQTLMRYGAGNFDWKEFVPRLKQKTFEEINEYGITFLKHIAEEIDENSPTFSDGVPKEG 1020  
|||||  
1021 LRIEDVLVRIALLILVQEKVKFVEDHPGKVPFSPRILERPGLRSGKIWKBEHDKIMIRA 1080  
|||||  
1021 LRIEDVLVRIALLILVQEKVKFVEDHPGKVPFSPRILERPGLRSGKIWKBEHDKIMIRA 1080  
|||||  
1081 VLKHGYGRWQAI VDDKEIGIQELICKELNFPHISLSAAEQAGLQGNNGSGSNPGAQTNQ 1140  
|||||  
1081 VLKHGYGRWQAI VDDKEIGIQELICKELNFPHISLSAAEQAGLQGNNGSGSNPGAQTNQ 1140  
|||||  
1141 NPGSVITGNNNASADGAQVNSMFYRDMQRRLLVEFVKRVLVLLLEKAMNIEYAEYYGLGG 1200  
|||||  
1141 NPGSVITGNNNASADGAQVNSMFYRDMQRRLLVEFVKRVLVLLLEKAMNIEYAEYYGLGG 1200  
|||||  
1201 SSSIPTEEPAEPKIAFTVGVSFIEVDDEMLDGLPKTDPITSEEIMGAADVNNQARVEIA 1260  
|||||  
1201 SSSIPTEEPAEPKIAFTVGVSFIEVDDEMLDGLPKTDPITSEEIMGAADVNNQARVEIA 1260  
|||||  
1261 QHYNQMCKLLDENARESVOAYVNNQPPSTKVNESFRALKSINGNINTILSITSDQKSHE 1320  
|||||  
1261 QHYNQMCKLLDENARESVOAYVNNQPPSTKVNESFRALKSINGNINTILSITSDQKSHE 1320  
|||||  
1321 DDTKPDNLNNVEMKDTAEETKPLRGVVDLNVVEGENIAEASGSVDVKMEEAKBEEKPKN 1380  
|||||  
1321 DDTKPDNLNNVEMKDTAEETKPLRGVVDLNVVEGENIAEASGSVDVKMEEAKBEEKPKN 1380  
|||||  
1381 MVVD 1384  
|||  
1381 MVVD 1384  
|||

RESULT 2  
B84645  
hypothetical protein At2g25170 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C;Accession: B84645  
R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.; euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J Nature 402, 761-768, 1999  
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A;Reference number: A84420; MUID:20083487; PMID:10617197  
A;Accession: B84645  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1359 <STO>  
A;Cross-references: GB:AE002093; NID:g6623882; PIDN:AAF19223.1; GSPDB:GN00139  
C;Genetics:  
A;Gene: At2g25170  
A;Map position: 2

Query Match 96.3%; Score 6921.5; DB 2; Length 1359;  
Best Local Similarity 96.6%; Pred. No. 0;  
Matches 1348; Conservative 0; Mismatches 0; Indels 47; Gaps 3;

QY 1 MSSLVERLIRSDRKPVYNLDDDDDFVPKKDRTFEQVEAIVRTDAKENACQACGESTN 60  
Db 1 MSSLVERLIRSDRKPVYNLDDDDDFVPKKDRTFEQVEAIVRTDAKENACQACGESTN 60  
QY 61 LVSCNTCTYAFHAKCLVPPLKQASVENWRCPECVSPNLNEIDKILDCENRPTKSSEQGSSD 120  
Db 61 LVSCNTCTYAFHAKCLVPPLKQASVENWRCPECVSPNLNEIDKILDCENRPTKSSEQGSSD 120  
QY 121 AEPKPIFVKQYLVKWKGLSYLHCSWVPEKQKAYKSNHRLKTRVNNFHRQMESFNNSD 180  
Db 121 AEPKPIFVKQYLVK-----VPEKEFKAYKSNHRLKTRVNNFHRQMESFNNSD 169  
QY 181 DFVAIRPEWTTVDRIILACREEDGELEYLVKYKELSYDECYWESESDISTFQNEIQRFKDV 240  
Db 170 DFVAIRPEWTTVDRIILACREEDGELEYLVKYKELSYDECYWESESDISTFQNEIQRFKDV 229  
QY 241 NSRTRRSKDVHKKRNPDRDFOQFDHTPEFLK-----GLLHPYQLEGLNFLRFSWS 289  
Db 230 NSRTRRSKDVHKKRNPDRDFOQFDHTPEFLKDVMIYLFPAIEGLLHPYQLEGLNFLRFSWS 289  
QY 290 KQTHVILADEMGLGKTIQSIALLASLFEENLI PHLVIAPLSTLRNWEREFATWAPQMNVV 349  
Db 290 KQTHVILADEMGLGKTIQSIALLASLFEENLI PHLVIAPLSTLRNWEREFATWAPQMNVV 349  
QY 350 MYFGTAQARAVIREHEFYLSKDQKKIKKKSQGISSESQKRIKFDVLLTSYEMINLDSA 409  
Db 350 MYFGTAQARAVIREHEFYLSKDQKKIKKKSQGISSESQKRIKFDVLLTSYEMINLDSA 409  
QY 410 VLKPIKWECEMIVDEGHRLLKNKDSKLFSSLTQYSSNNHRI LLTGTPLQNNLDELFMHFLD 469  
Db 410 VLKPIKWECEMIVDEGHRLLKNKDSKLFSSLTQYSSNNHRI LLTGTPLQNNLDELFMHFLD 469  
QY 470 AGKFGSLEEFQEEFKDINQEEQISRLHKMLAPHLLRRVKKDVNKMPPPKKELILRVDLSS 529  
Db 470 AGKFGSLEEFQEEFKDINQEEQISRLHKMLAPHLLRRVKKDVNKMPPPKKELILRVDLSS 529  
QY 530 LQKEYYKAIFTRNYQVLTCKGGAQISLNNIMMELRKVCCHPYMLEGVEPVIIHDANEAFKQ 589  
Db 530 LQKEYYKAIFTRNYQVLTCKGGAQISLNNIMMELRKVCCHPYMLEGVEPVIIHDANEAFKQ 589  
QY 590 LLESCGKLQLLDKMMVKLKEQGHRLVITYTQFQHMLDLEDYCTHKKWQYERIDGKVGGA 649  
Db 590 LLESCGKLQLLDKMMVKLKEQGHRLVITYTQFQHMLDLEDYCTHKKWQYERIDGKVGGA 649  
QY 650 RQIRIDRFNAKNSKFCFLSTRAGGLGINLATADTVIIYDSWNPHADLQAMARAHRLG 709  
Db 650 RQIRIDRFNAKNSKFCFLSTRAGGLGINLATADTVIIYDSWNPHADLQAMARAHRLG 709  
QY 710 QTNKVMYIRLINRGTIERMMQLTKKKMWLEHLVVGKLTQNTINQEEELDDIIRYGSKELF 769  
Db 710 QTNKVMYIRLINRGTIERMMQLTKKKMWLEHLVVGKLTQNTINQEEELDDIIRYGSKELF 769  
QY 770 ASEDDEAGKSGKIHYYDDAAIDKLLDRDLVEAEVSVDDDEENGFLKAFKVANFEYIDENE 829  
Db 770 ASEDDEAGKSGKIHYYDDAAIDKLLDRDLVEAEVSVDDDEENGFLKAFKVANFEYIDENE 829

QY	830	AAALEAQRVAAESKSSAGNSDRASYWEELLKDKFELHQAEEALNALGKRKRKQLVSIIE	889
Db	830	AAALEAQRVAAESKSSAGNSDRASYWEELLKDKFELHQAEEALNALGKRKRKQLVSIIE	889
QY	890	DDLAGLEDVSSDGDESYEAESTDGEAAGQGVQTRRRPYRRKGRDNLEPTPLMEGGRSFR	949
Db	890	DDLAGLEDVSSDGDESYEAESTDGEAAGQGVQTRRRPYRRKGRDNLEPTPLMEGGRSFR	949
QY	950	VLGFNQSORAIFVQTLNRYGAGNFDWKFEVPRLKQKTFEEINEYGILFLKHIAEEIDENS	1009
Db	950	VLGFNQSORAIFVQTLNR-----YGILFLKHIAEEIDENS	984
QY	1010	PTFSDGVPKEGLRIEDVLRIALLILVQEKVKFVEDHPGKPVFPISRILERPGLRSGKIW	1069
Db	985	PTFSDGVPKEGLRIEDVLRIALLILVQEKVKFVEDHPGKPVFPISRILERPGLRSGKIW	1044
QY	1070	KEEHDKIMIRAVLKHGYGRWQAIVDDKELGIGIQLICKELNFPFHISLSAAEQAGLQGNGS	1129
Db	1045	KEEHDKIMIRAVLKHGYGRWQAIVDDKELGIGIQLICKELNFPFHISLSAAEQAGLQGNGS	1104
QY	1130	GGSNPGAQTNQNPGSVITGNNNASADGAQVNSMFYYRDMQRRLLVEFVKRRVLLLEKAMNY	1189
Db	1105	GGSNPGAQTNQNPGSVITGNNNASADGAQVNSMFYYRDMQRRLLVEFVKRRVLLLEKAMNY	1164
QY	1190	EYAEYYGLGSSSIPTEEPAEPKIADTVGSFIEVDDEMLDGLPKTDPITSEEMGAA	1249
Db	1165	EYAEYYGLGSSSIPTEEPAEPKIADTVGSFIEVDDEMLDGLPKTDPITSEEMGAA	1224
QY	1250	VDNQARVEIAQHYNQMKLLDENARESVAQVYNNQPPSTKVNESFRALKSINGNINTIL	1309
Db	1225	VDNQARVEIAQHYNQMKLLDENARESVAQVYNNQPPSTKVNESFRALKSINGNINTIL	1284
QY	1310	SITSQSKSHEDDTKPDINNVMKDTAEETKPLRGVVVDLNVVEGEENIABASGSVDVKM	1369
Db	1285	SITSQSKSHEDDTKPDINNVMKDTAEETKPLRGVVVDLNVVEGEENIABASGSVDVKM	1344
QY	1370	EEAKEEEKPNMVVD	1384
Db	1345	EEAKEEEKPNMVVD	1359
RESULT 3			
T06312			
hypothetical protein F11C18.100 - Arabidopsis thaliana			
C;Species: Arabidopsis thaliana (mouse-ear cress)			
C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 22-Oct-1999			
C;Accession: T06312			
R;Bevan, M.; Terryn, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.; De			
ewes, H.W.; Mayer, K.F.X.; Schueller, C.			
submitted to the Protein Sequence Database, April 1999			
A;Reference number: Z15589			
A;Accession: T06312			
A;Molecule type: DNA			
A;Residues: 1-1067 <BEV>			
A;Cross-references: EMBL:AL049607; GSPDB:GN00062; ATSP:F11C18.100			
A;Experimental source: cultivar Columbia; BAC clone F11C18			
C;Genetics:			
A;Gene: ATSP:F11C18.100			
A;Map position: 4			
A;Introns: 40/3; 93/2; 151/2; 223/1; 255/1; 300/3; 326/1; 374/3; 407/2; 453/3; 510/2; 53			
Query Match			
Best Local Similarity 43.3%; Score 3114; DB 2; Length 1067;			
Matches 673; Conservative 155; Mismatches 205; Indels 332; Gaps 25;			
QY	1	MSSLVERLRIRSDRPVYNLDDSDDDFVPPKDKRTFEQVEAIVRTDAKENACQACGESTN	60
Db	1	MANLLQRLRRRTGPKPDYIEDKLDE-----YIREEQVEETGGSNQDC-----	42
QY	61	LVSCNTCTYAFHAKCLVPPLKQASVENWRCPECVPLNEIDKILDCEMPKTSSEQSSD	120
Db	43	-----PLGEIEKILDREWPTASNNPSSD	67

QY	121	-AEPKPIFVKQYLVKWKGLSYLHCSWVPEKBFQKAYKSN-H-RLKTRVNNFHRQMESF--	175
Db	68	NGTPTLVVVKQYLVKWKGLSYLHCSWVPEQEFKAYKSHPHLKJLKRVTRENAAMDVFA	127
QY	176	NNSEDDFVAIRPEWTTVDRIILACREEDGELEYLVKYKELSYDECYWESESDISTFQNEIQ	235
Db	128	ENGAHEFAIRPEWKTVDRIIACREGDDGGEYLVKYKELSYRNSYWESESDISDFQNEIQ	187
QY	236	RFKDVNSRTRRSKDVDRKRNPRDFQQFDHTPEFLKGLLHPYQLEGLNFRFSWSKQTHVI	295
Db	188	RFKDINSSRRDKYVENERNREEFKQFDLTPEFLTGTLHTYQLEGLNFRYSWSKKTNVI	247
QY	296	LADEMGLGKTIQSIALLASLFEENLIPLHVIAPLSTLRNWEREFATWAPQNVVVMYFGTA	355
Db	248	LADEMGLGKTIQSIAPLASLFEENLSPHLVVAPLSTIRNWEREFATWAPHNVVVMYTGDS	307
QY	356	QARAVIREHEFYLSKDQKKIKKKKSGQISSESQKRIKFDVLLTSYEMINLDSAVLKPIK	415
Db	308	EARDVIWEHEFYFSEGRK-----	325
QY	416	WECMIVDEGHRLLKNKDSKLFSSLTQYSSNHRILLTGTPLQNNLDELFLMHFLDAGKFGS	475
Db	326	-----	325
QY	476	LEEFQEEFKDINQEEQISRLHKMLAPHLRRVRVKDVMKO-MPPKKELILRVDLSSLQKEY	534
Db	326	-----RLKDVLDKQVPPKKELILRVDMSQQKEV	355
QY	535	YKAIFTRNYQVLTCKGGAQISLNNIMMELRKVCCHPYMLEGVEPVIHDAEAFKQLLESC	594
Db	356	YKAVITNNYQVLTCKRDAKIS--NVLMKLRQVCSHPYLLPDFEPRFEDANEAFTKLEAS	413
QY	595	GKLQLLDKMMVKLKEQCHRVLITYTQFQHMULDLEDYCTHKKQYERIDGKVGGAEQIRI	654
Db	414	GKLQLLDKMMVKLKEQCHRVLITYTQFQHTLYLLEDYFTFKNNWYERIDGKISGPERQVRI	473
QY	655	DRFNAKNSNKFCLLSTRAGGLGINLATADTVIIYDSDNPHADLOAMARAHRLGQTNKV	714
Db	474	DRFNAENSNRFCFLSTRAGGINLATADTVIIYDSDNPHADLOAMARVHRLGQTNKV	533
QY	715	MIYRLINRGTEERMMLTKKQVLEHLVVGKLTQNNINQEEELDDIIRYGSKELFASEDD	774
Db	534	MIYRLIHKGTVEERMMEITKNKMLLEHLVVGK--OHLQCELDLDDIIRYGSKELFSEEND	590
QY	775	EAGSGKIHVDDAAIDKLLDRDLVEAEVSVDDDEENGFLKAFKANFEYI-DENEAAL	833
Db	591	EAGSGKIHVDDAAIEQLLDRNHVDAVEVSLDDEETDFLKNFKVASFEYVDDENEAAL	650
QY	834	EAQVAAESKSSAGNSDRASYWEELLKDKFELHQAEEALNALGKRKRKQLVSIIEEDDLA	893
Db	651	E-EAQAIENSSVRNADRTSHWKDLLDKYEVQQAEEALSALGKRKRNGKQ-VMYAEDDLD	708
QY	894	GLEDVSSDGDE-----SYEAESTDCEAAGQ-GVQTGRRPYRRKGRDNLEPTPLME	942
Db	709	GLEEISDEDEYCLDDLKVTSDDEEEADEPEAARQKPRVTVTRPYRKARDNSEEIPLME	768
QY	943	GEGRSFRVLGFNQSORAIFVQTLNRYGAGNFDWKFEVPRLKQKTFEEINEYGILFLKHIA	1002
Db	769	GEGRYLMVLGFNETERDIFLRTFKR-----YGILFLKHIA	803
QY	1003	EEIDENSPTE-----SDGVPKEGLRIEDVLRIALLILVQEKVKFVEDHPGKPVFPS	1054
Db	804	ENPTDNSTNFKVITAMVYADGVPKEGISSDELLVSMTFMMLVKEKQCQFLDNHPTAPVFN	863
QY	1055	RILERFPGRLSGKINWKEEHDKIMIRAVLKHGYGRWQAIVDDKELGIGIQLICKELN--FPH	1112
Db	864	YVISKY-NLRNGAFSKEEHDRIILIPAVSKHGYGRWVAIVEDDEEIGFQEVACKDLNIPFPP	922
QY	1113	ISLSAAEQAGLQGNGSGSNPGAQTNQNPGSVITGNNNASADGAQVNSMFYYRDMQRRLL	1172
Db	923	DTKSA-----RKRI	931



QY 1173 VEFVKRVLLEKAMNVEYAEYYGLGSSSIPTPEEPAEKIADTVGVSPIEVDDMLD 1232  
Db 932 CDHVGKRVKMEDAIKVEAEKIL-----AEQAKAETKGTSPDAEKEML- 976  
QY 1233 GLPKTDPITSEEIMGAADVNNQARVEIAQHVNNQCKLLDENARESVOAYVNNQPPSTKVN 1292  
Db 977 ---KNDPITSKKSATSADVNNKQGRVEMAQSYDQS---VNEKSGESFQTYLDIQPLNRMPR 1030  
QY 1293 ESFRALKSINGNINTILSITSDQSKSHEDDTKPDNNVEMKDTAE 1337  
Db 1031 ESFKPLEPINEEISTRLSVGTD-----HDVEM-DAAD 1061

RESULT 4  
T34239  
hypothetical protein F26F12.7 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C;Accession: T34239  
R;Wilson, R.; Bentley, D.; Gattung, S.  
submitted to the EMBL Data Library, April 1996  
A;Description: The sequence of C. elegans cosmid F26F12.  
A;Reference number: Z21493  
A;Accession: T34239  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1829 <WIL>  
A;Cross-references: EMBL:U55373; PIDN:AAC25894.1; GSPDB:GN00023; CESP:F26F12.7  
A;Experimental source: strain Bristol N2; clone F26F12  
C;Genetics:  
A;Gene: CESP:F26F12.7  
A;Map position: 5  
A;Introns: 110/3; 441/3; 801/2; 1244/3; 1693/2; 1784/1

Query Match 28.5%; Score 2048.5; DB 2; Length 1829;  
Best Local Similarity 32.9%; Pred. No. 9.6e-96;  
Matches 518; Conservative 253; Mismatches 474; Indels 331; Gaps 44;  
QY 40 EAIVRTDAKEN--ACQACGESTNLVSCNTCTYAFHAKCLVPLPKDA-SVENWRCPEC--V 94  
Db 306 EVVKEEPAQNDEFCKICKETENLLCDSVCVCFHAYCICDPPLTEVPKEETWSCRCETV 365  
QY 95 SPUNEIDKILDCMR-----PTKSSEQGSSD-----AEPKPIFVKQYLVKWKGLSYL 141  
Db 366 KPEHKIEKILCWRWKEIPYPEPLEAGKEASSDDAMLKPPRKVEPRREREFFVKWKYLSYW 425  
QY 142 HCSWVPEKEFQKAYKSNHRLLKTRVNN-----FHRQMESFNNSBDD-----FVAIR 186  
Db 426 QCSWVSEMILLEVHFRMLILLVWRKNDSAPPEFEESVTSRHSNDNDPYKLRRFYQYGIK 485  
QY 187 PEWTTVDRIACRE-EDGELEYLVKVKELSYDECYWE-SESDISTFQNEIQRF----- 237  
Db 486 PEWQIHRINIHQSYAKSQDYLVKWKELSYDQATWERDSDNIANYEEAIIKYWOHRESK 545  
QY 238 --KDVNSRTRR-----SKDQVHKNRPD----FQQFDHTPEFLK---GLLHPY 276  
Db 546 LNEIDPKNVQKMIKHKREAKGLPPKEDEKKKKREKIDIRKKYEVQPDYVTTETGKGLHPY 605  
QY 277 QLEGNPLRFSWSKQTHVILADEMGLGKTIQSIALLASLFEENLI--PHLVIAPLSTLRN 334  
Db 606 QLEGNWLHRHCWSNGTDAILADEMGLGKTVQSLTFLYSLMKEGHCKGPFLLIAAPLSTIIN 665  
QY 335 WEREATWAPQMNVMVYFGTAQARAVIREHEFYLSKDQKKIKKKKSGQISSSESQKRIKF 394  
Db 666 WEREABQWCPDFYVVTYVGLRDARVVLREHEFSFVEGAVRSGPKASKMKTEN-----MKF 721  
QY 395 DVLLTSYEMINLDSAVLKPIKWCMIVDEGHRLLKNKOSKLFSSLTQYSSNHRILLTGTP 454  
Db 722 HVLLTSYETINMDKTILSSIEWGALVVDGAHRLKNNQSLFFFKNLNEYTIHYRVLLTGTP 781  
QY 455 QNNLDELFMHFLDAGKFGSLEEFQEEFKDINQEEQISRLHKWLAPHLLRRVKKDVMD 514  
Db 782 QNNLEELFHLNFLSKERFNQLEAFTAEFNEISKEDQIEKLHNLGLPHMLRRLKADVL 841

QY 515 MPPKKELILRVDLSSLQKEYYKAIFTRNYQVL-TKKGGAQISLNNIMMELRKVCCHPYML 573  
Db 842 MPSKSELIVRVELSAMQKKWYKNILTRNFDALNVKNGGTQMSLMNVLMELKCCNHPYLF 901  
QY 574 EGVEPVIHDANEAFKQ-----LLESCGKLQLLDKMMVKLKEQGHRVLIYTFQFHML 624  
Db 902 VKAE-----LEAPKEKNGMYEGTALIKNSGKFVLLQKMLRKLKDGGRVLIIFSQMT 955  
QY 625 DLLEDYCTHKKWQYERIDGKVGAERQIRIDRFNAKNSKFCFLLSSTRAGGLGINLATAD 684  
Db 956 DIMEDLCEYEGYRERIDGSIHQMRQDAIDRYNAPGAQQFIFLLSTRAGGLGINLATAD 1015  
QY 685 TVIIYDSWNPHADLQAMARHRLGQTNKVMYIRLINRGTIERMMQLTKKKMVLHLLV 744  
Db 1016 TVIIYDSWNPNDIQAFSRAHRLGQHKVMYIRFVTKSVEEKITSVAKKKMLLNHLV 1075  
QY 745 ---GKLTQINQEEELDDIIRYGSKELFASEDDDEAGKSG-----KIHVDDAAIDK 791  
Db 1076 RAGLGKKEGKMTSKTELDVLRWGTEELFSEDLDAEAGESEKKGAAQAQIIVWDDAAVDA 1135  
QY 792 LLDRLVAAEEVSVDDDE---ENGFLKAFKAVANFEYIDENEAALAEQRAAESKSSAG 847  
Db 1136 LLDRSNKEETPAGEDGEEKAEWQNEYLSSFVASYQ-TKETEGQEEEEEEETEVIKEDEK 1194  
QY 848 NSDRASYWEELLKDKFELHQAEELNALGKRKRSRKQLVSIIEED---DLAGLEDVSSDGD 904  
Db 1195 EPD-PDYWEKLLKHHEQDREIELOKLGKGRVVKQINYASENMGTDWSKQNTQDQDDDD 1253  
QY 905 SYEAESTDGEAAGQVQTRRPY---RRKGRDNLEPTPLMEGEGRSFRVLGENQSQR 961  
Db 1254 N---ESYRGSDNGDLNSDEDDYDEKKRRRDEEKMPPLMAKVNGQVEILGFENPRQRKAF 1310  
QY 962 VQTLMYGAGNFD---WKEFVPRLKQKTFEINEYGIILFKHIAEEIDENSPTFS 1016  
Db 1311 YGAVMRWGMPPQDSHQSQW--LVRLRNKSEKVFRAVASLFRHLCEPGADGHDTFNDGV 1368  
QY 1017 PKEGLRIEDVLRIALLILVQEKVKFVEDHPG----- 1048  
Db 1369 PREGLNHQHVLGRIGLLSLVRKQVEFEQYNGQWSMPEIQDEVLAKAANGSAQSSSRSTP 1428  
QY 1049 ----- 1048  
Db 1429 KPKEEPMKEKEDATETVNGATSEPATDAESEQNAPVDEPMDTDEAKEPEPIETEK 1488  
QY 1049 ---KPVPFSRILE-RPPGL-----RSGK---IWKEEHDKIMIRAVLKHGGRW 1089  
Db 1489 PRAARPSFKNICDGGFTLHSLWANEKVARNGKEYEIWYRRHDYWLLAGVVVHGYGRF 1548  
QY 1090 QA-----IVDD-----KELGI-----QELICKE----- 1107  
Db 1549 QANFNDIINDPRFSVLNPPFKEVGAEATGSDIKAKPMQRRFKLIEQSLVIEEQLRAAHA 1608  
QY 1108 -----LNFPHISLAAEQAGLQNGSGSNPQAQTNQNPNGSVITGNNNASA 1154  
Db 1609 NRHLQPDNVGPLAQRFAADLENIAESQANIAKESAGNRNANAVLHKTLLVQLDEILSDMKA 1668  
QY 1155 DGAQVNSMFYRDMQRLVFEVFKRVLLEKAMNVEYAEYYGLGSSSIPTPEEPEAPPK 1214  
Db 1669 DVSRLPSTF-----TQLATVTERLNMTERQILSRLLTKDEDAIANRSLVPPPPFFVTPI 1722  
QY 1215 I---ADTVGVSFIEVDDEML---DGLPKTDPITSEEIMGAADVNNQARVEIAQHVNNQMC 1267  
Db 1723 LRQQMDGIQPKFAALYSKFMSENGERMEEDPEVAEEBEGVKQEP----- 1767  
QY 1268 KLLDENARESVOAYVNNQPP--STKVNESFRALKSINGNINTILSITSDQSKSHEDDTKP 1325  
Db 1768 ---DDETQDSAEA-----PPVLSAEVN-----SDSDNDVPSTAAAAVSSSE--TAA 1808  
QY 1326 DLNNVEMKDTAEETKP 1341  
Db 1809 DAEPAEADQAPTDEP 1824







QY 119 --SDAEPKPIFVK-----QYLVKWKGLSYLHCSWVPEKEFQKAYKSNHRLKTRVNFH 169  
Db 295 VEADGDPNAGFERNKEPGDIOYLKWKGWSHIHTWETEETLK---QQNVRGMKKLDNYK 351  
QY 170 RQ-----MESFNSEDFFVAIRPEWTTVDRIACREED---GELEYLVKY 211  
Db 352 KKDQETKRWLNKASPEDVEYINCQOQELTDDLHKQYQIVERIIAHSNKQSAAGLPDYCKW 411  
QY 212 KELSDECYWESESDIS-TFQNEIQRFKDVN-SRTRRSKDVD-HKRNPRDFQQFDHTPEF 268  
Db 412 QGLPYSECSWEDGALISKKFQTCIDEYFSRNQSKTTPFKDCKVLKQRP-RFVALKKQPSY 470  
QY 269 LKG---LLHPYQLEGLNFLRFSWSKQTHVILADEMGLGKTIQSIALILASLFEENLI--P 322  
Db 471 IGGHEGLELRDYQLNGLNWLASHWCKNGNSCILADEMGLGKTIQITISFLNYLFHEHQLYGP 530  
QY 323 HLVIAPLSTLRNWEREFATWAPQMVVMYFGTAQARAVIREHEFYLSKDQKIKKKKSGQ 382  
Db 531 FLLVPLSTLTSWQREIQTWASQMNNAVYLGIDINSRNMIRTHEW-----574  
QY 383 ISSESQKRIKFDVLLTSYEMINLDSAVLPKPKWECMIVDEGHRLLKNKDSKLFSSLTQYS 442  
Db 575 --MHPQTKRLKENILLTTYEILLKOKAFLGGLNWAFIGVDEAHLKNDSDSLLYKTLIDFK 632  
QY 443 SNHRILLTGTPLONNLDELFLMHFLDAGKFGSLEEFQEEFKDINQEQISRLHKMLAPH 502  
Db 633 SNHRLITGTPLQNSLKELWSLLHFIMPEKFSWEDFEEHGK-GREYGYASLHKELEPF 691  
QY 503 LLRRVKKDVMDPMPKKEILIRVDLSSLOKEYKKAIFTRNYQVLT-KGGAQISLNNIMM 561  
Db 692 LLRRVKKDVESLPAKVEQILRMEMSALQKYKWLTRNYKALSKGSGSTSGFLNIMM 751  
QY 562 ELRKVCCHPYMLEGVE-PVIHDANEAFKQLLESCGKLQLLDKMMVKLKEQCHRVLIYTQF 620  
Db 752 ELKKCCNHCYLKPPDNNEFYNKQEALQHLIRSSGKLILLDKLLIRLRERGNRVLIFSOM 811  
QY 621 QHMLDLLEDYCTHKKWQYERIDGKVGGAERQIRIDRFNAKNSKFCFLLSTRAGGLGINL 680  
Db 812 VRMLDILAEYLKYRQFPFQRLDGSIKGELRKOALDHFNAEGSEDFCFLSTRAGGLGINL 871  
QY 681 ATADTVIYDSWNPHADLOAMARAHRLGQTNKVMYRLINRGTIERMMQLTKKMVLE 740  
Db 872 ASADTVIFDSWNPQNDLQAQARAHRIQGGKQVNIYRLVTKGSVEEDILERAQKMWLD 931  
QY 741 HLVVGKLKTQN-----INQEELDDIIRYGSKELFASEDDEAGSKGIHYD 785  
Db 932 HLVIQRMDDTTGKTVLHTGSAPSSSTPFNKEELSAILKFGAEELFKEPEGEHEQEP-----Q 986  
QY 786 DAAIDKLLDRDLVEABE---VSVDDDEENGFLKAFKVANFEYIDENEAALAEQORVAES 842  
Db 987 EMDIDEILKRAETHENEPGLSVGDE---LLSQFKVANFSNMDEDD-IELEPER---NS 1038  
QY 843 KSSAGNSDRASVWEEL-----LKDKFELHQAEEELNALGKRKRSRKQLVSIEDDLG 894  
Db 1039 KN-----WEEIIPEEQRRRLEEEERQKELEIYMLPRMRNCAQ-----1077  
QY 895 LEDVSSDGDESYAEASTDGEAAGQGVQGRPYRKRGRDNLEPTPLMEGEGRSFRVLGFN 954  
Db 1078 ---ISFNGSEGRRSRRYSGSDSDSISERKPKKRGRPTIPRE-----NIKGS 1125  
QY 955 QSQRAIFVQTLMRYGAGNFDWKEFVPR---LKQKTFEEINEYGILFLKHIAEEIDENS-- 1009  
Db 1126 DAEIRRFIKSYKKFG-GPLERLDAIARDAELVDKSETDLRRLGELVHNGCVKALKDSSSG 1184  
QY 1010 -----PTFSDGVPKEGLRIEDVLRIALLILVQEKVKFVEDHPKPVFPSPRI 1056  
Db 1185 TERAGGRGLKVKGPTF-----RISGVQVNAKLVIHAHEDEL--IPLHKSIPSDPE-- 1231  
QY 1057 LERFPGLRSGKI-----WKEEHDKIMIRAVLKHGYGRWQAIVDDKELGIGELI 1104  
Db 1232 -ER-----KQYTIPTCHTKAAHFIDWKGEDDSNLLIGIYEYGSWEMIKMDPDLSLTHKI 1286

QY 1105 -----CKELNFPH---ISLSAAEQAGLQGN--GSGGSNPGAQTNQNPQSVITGN 1149  
Db 1287 LPDDPDKKPQAKQLQTRADYLIKLSRDLAKREAQRLCGAGS-----1329  
QY 1150 NNASADGAQVNSMFYYRDMQRRLVEFVKRVLLEKAMNIEAEEYYGLGSSSIPTEEP 1209  
Db 1330 -----KRRKTRAKKSKAM-----KSIKVKEEI--KSDSSPLPSEKS 1363  
QY 1210 EAEPKIADTVGVSFIEVDDDEMLDGLPKTDPITSEIM-----GAAVDNNQARVEI 1259  
Db 1364 D-----EDDDKLNDSKPESKDRSKSVSDAPVHITASGEPPVPIAESEEL 1409  
QY 1260 AQHYNQMCK 1268  
Db 1410 DQKTFISICK 1418  
RESULT 8  
C84507  
hypothetical protein At2g13370 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 23-Dec-2002  
C:Accession: C84507  
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A;Reference number: A84420; MUID:20083487; PMID:10617197  
A;Accession: C84507  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1738 <STO>  
A;Cross-references: GB:AE002093; NID:G4733988; PIDN:AAD28668.1; GSPDB:GN00139  
C;Genetics:  
A;Gene: At2g13370  
A;Map position: 2  
C;Superfamily: chromodomain helicase CHD1; chromobox homology  
Query Match 21.8%; Score 1565.5; DB 2; Length 1738;  
Best Local Similarity 30.7%; Pred. No. 2.7e-71;  
Matches 462; Conservative 237; Mismatches 485; Indels 321; Gaps 50;  
QY 11 RSDRKPVYNLDDSDDDDFVPKKDRTFEQVEAIVR--TDAKENACQACGESTNLV-----SC 64  
Db 329 RQKRKTSYQDDSEED---SENDNDEGRSLARRGTTLRQNN---GRSTNTIGQSSEV 380  
QY 65 NTCTYAFHAKCLVPPLKDASVENWRCPECVSPLNE-----IDKILDCEMRPTKS 113  
Db 381 RSSTRSVRKVSVESEDSIEDDGK-----NRKNQKDDIEEEDADVIEKVLWHQLKGMGE 435  
QY 114 SEQSSSDAEPKPIFVKQ-----YLVKWKGLSYLHCSWVPEKEFQK-----153  
Db 436 DVQ-TNNKSTVPVLVSQLFDTPEPDWNEFELIKWKQSHLHCQWKTLSDQLNSLGFKKVL 494  
QY 154 --AYKSNHRLKTRVNNFHRQMESFNNSDEDDFVAIRPEWTTVDRIACR-BEDG---ELE 206  
Db 495 NYTKKVTEIRYRTALSRREEIEVNDVSKEMDLDIIKQNSQVERIADRISKDGLGDVVPE 554  
QY 207 YLVKYKELSYDECYWESESDISTFQNEIQRFKDVN-SRTRRSKDVDHXR-----254  
Db 555 YLVKWQGLSYAEATWEKVDVIAFAQVAIDYKAREVSIQVQGMVEQQRTKKGKENSFSN 614  
QY 255 -----NPRDFQQFDHTPEFL-KGLLHPYQLEGLNFLRFSWSKQTHVILADEMGLGKTI 306  
Db 615 AELWLLFSVASLRKLDQEPPEWLIGGTLRDYQLEGLNFLVNSWLNDTNVILADEMGLGKTV 674  
QY 307 QSIALLASLFEENLI--PHLVIAPLSTLRNWEREFATWAPQMVVMYFGTAQARAVIREH 364  
Db 675 QSVMLGFLQNTQQIIPGFVLVVPLSTLANWAKEFRKWLPGMNIIVYVGTASREV----730  
QY 365 EFYLSKDQKKIKKKKSGQISSSESQKRIKFDVLLTSYEMINLDSAVLPKPKWECMIVDEG 424



Db	749	LRRFPSQRLDGSMRADLRKQALDHYNAPGSTDFAPLLSTRAGGLGINLATADTVIIFDSD	808
Qy	693	WNPHADLQAMARHLGQTNKVMYIRLINRGTIERMMQLTKKMVLEHLVVGKLKTQN-	751
Db	809	WNPQNDLQAMSRAHRIGQTKTVNIYRLVTKGSVEEIVERAKRKLVDLHLVIQRMDDTTGK	868
Qy	752	-----INQEELDDIIRYGSKELFASEDDDEAGSKGIHYDDAIDKLLDRDL	797
Db	869	TVLSKNATASGVFPDKQELSAILKFGAVELFKEGEE-----QEPEVD--IDRIL	918
Qy	798	VEAEVSVDD--EENGFLKAFKVFANFEYIDENEAALAEQRAAEKSSAGNSDRASYW	855
Db	919	MGAETRAEEVVMKENELSSFKYANFAIDEEKDIAATDE-----W	960
Qy	856	BELLKDK-----FELHQAEEELNALGKRKRKQ-LVSIIEEDDLAGLEDVSSDGDESYAE	909
Db	961	AAIIPEDRNRILEEERMKELAEMLNAPRQKQIPQVVEDD-----DGDDEEED	1011
Qy	910	STDGEAAGQVQIGRRPYRKGRDNLEPTPLMEGEGRSFVLGFNQSORAIFVQTLMRYG	969
Db	1012	-----DTGKKK-KKKA VGNF-TIPEIKRFIKSFRK-----FSMPLNRLE	1048
Qy	970	--AGNFDWKEFVPRLKQKTFFEEINEYIGILFLKHAEEIDENSPTFSGDGVPKEGLRIEDVL	1027
Db	1049	EIAQDAELEEHSTDEMCKLVESLSE---ACKKADEFDSNEKNGDAGAASEKKD---	1100
Qy	1028	VRIALLILVQEKVF-----VEDHPGKVPFPPSRILE-RFPGLRSGKI-----	1068
Db	1101	-----IERKFKFHTCDVNLKQIERSHAELKPLHLKSEETKTSFKPPANAKLQXGW	1152
Qy	1069	---WKEEHDKIMTRAVLKHGYGRWQAI VDDKELGIGELI-----	1104
Db	1153	DVDWSRPDDSAALLGVWKYGYGSWEAIKMDPTILGLADKIFIKDKTKPKQGNLQVRVDYL	1212
Qy	1105	CKELNFPHISLSAAEQAGLQGGSGGNPGAQTNQNPQSVITGNNNASADGAQVNSMFY	1164
Db	1213	LKLMKDKVKTTTEKERRKADDPVG-----PEKKRHTNNVPQEGEKKK---	1259
Qy	1165	YRDMQRRLVEFVKRVLLL--EKAMNYEYAEYYG--LGGSSSIP-----	1205
Db	1260	EKKEKNSSSLKDQLALLSIDKSL-----YGALEDSSAKPFLECVKLCMPVHKYM	1310
Qy	1206	---TEEPEA-----EPKIADTVGVSFIEVDDMDGLPKTDPITSEEINGAAVDNNQARV	1257
Db	1311	KKLKEAQEAQNQDAEAKYLTRLGDSFLENLETLIKKKPKTN-----	1351
Qy	1258	EIAQHYNQ---MCKLLDENARESVOAYVNNPPSTKVNESFRALKSINGINTILSITS	1313
Db	1352	IRKWNYLWIFLCKFTLREPGEADRY-----RSITS	1383
Qy	1314	DQSKSH-----EDDTKPDLNNVEMKDTAEETKPLRGGVVDLNVVEGEENIAEASG-SV	1365
Db	1384	DKHKNHHHHHKKSEKPK---EAKDHKE-----RDREKORERNRGERMDHGEGETSK	1434
Qy	1366	DVKMEEAKEEEK	1377
Db	1435	DHREHHKDHK	1446
RESULT 10			
Tl3944			
chromodomain helicase - fruit fly (Drosophila melanogaster)			
C;Species: Drosophila melanogaster			
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Dec-2002			
C;Accession: Tl3944			
R;Stokes, D.G.; Tartof, K.D.; Perry, R.P.			
Proc. Natl. Acad. Sci. U.S.A. 93, 7137-7142, 1996			
A;Title: CHD1 is concentrated in interbands and puffed regions of Drosophila polytene ch			
A;Reference number: Z17823; MUID:96293489; PMID:8692958			
A;Accession: Tl3944			
A;Status: preliminary; translated from GB/EMBL/DBJ			
A;Molecule type: mRNA			
A;Residues: 1-1883 <STO>			

A;Cross-references: EMBL:L77907; NID:gl448982; PID:gl448983; PIDN:AAC37264.1			
C;Genetics:			
A;Gene: CHD-1			
A;Cross-references: FlyBase:FBgn0016132			
C;Superfamily: chromodomain helicase CHD1; chromobox homology			
C;Keywords: DNA binding			
Query Match 21.2%; Score 1520.5; DB 2; Length 1883;			
Best Local Similarity 29.1%; Pred. No. 5.9e-69;			
Matches 429; Conservative 253; Mismatches 459; Indels 333; Gaps 49;			
Qy	6	ERLIRSDRKPVYNLDDSDDDFVPKKDRTFEQVEAIVRTDAKENACQACGESTNLV---	62
Db	248	QRRRV---KP-FSSEDSDDDD-ASKRCATRKRKGAAVSYKEASEDEAT---DSEDLLEFE	298
Qy	63	--SCNTCTYAFHAKCLVPPLKDasVENWRCPECVSPLNIDKIL-----DCEMRPTK-	112
Db	299	YDEQAATAATAE-----EKEC-----ETIERILAQRAKRGCTGNQTTI	340
Qy	113	-----SSEQSSDAEPKPIFVKQYLWKWGLSYLHCSWPE-----	148
Db	341	YAIENGFDPHAGFDEKQTPDAETE-----AQFLIKWKGWSYIHNWSEATLRDMKAKGM	396
Qy	149	-----KEFQKAYKSNHRLKTRVNNF-----HRQMESFNNSDEDDFVAIRPEWTTVD	193
Db	397	KKLDNFIKKEQAYWRRYAGPEDIDYFECQLELQHELLKSYNN-----VD	442
Qy	194	RILA--CREEDGELEYLVKYKELSYDECYWESES-DISTFQNEIQRFKDVNSRTRRSKDV	250
Db	443	RIIAKGSKPDGTEEYLCKWQSLPYAESTWEDAALVLRKWQRCAEQFND-----RESSKC	497
Qy	251	DKRNP-----DFQQFDHTPEFL-KGL-LHPYQLEGLNFLRFSWSKQTHVILADEMGL	302
Db	498	TPSRHCRVTKYRPFKFSRIKNOPEFLSSGLTRDYQMDGLNWLHLSWCKENSVILADEMGL	557
Qy	303	GKTIQSIALLASLFEENLI--PHLVIAPLSTLRNWEREFATWAPQNVVMYFGTAQARAV	360
Db	558	GKTIQITICFLYSLFKIHLYGPFLCVVPLSTMTAWQREFDLWAPDMNVVYTLGDIKSREL	617
Qy	361	IREHEFYLSKDQKKIKKKSGQISSKQKRIKFDVLLTSYEMINLDSAVLKPIKWECMI	420
Db	618	IQYEWQF-----ESSKRLKFCILTTYEIVLKDQFLGTLQWAALL	659
Qy	421	VDEGRLKNKDSKLFSSLTQYSSNHRILLTGTPLQNNLDELFLMHFLDAGKFGSLEEFQ	480
Db	660	VDEAHLKNDDSLLYKSLKEFTDNRHLITGTPLQNSKELWALLHFMFMPDKFDTWENFE	719
Qy	481	EEFKDINQEEQISRLHKMLAPHLRRVKVDVMDMPKPKELILRVDLSLQKEYYKAIFT	540
Db	720	VQHGNA-EDKGHTRLHQOLEPYILRRVKDVEKSLPAKVEQILRVENTSLQKQYKWLIT	778
Qy	541	RNYQVLTK-KGGAQISLNNIMMELRKVCCHPYMLEGVEPVH--DANEAFKQLLESCGL	597
Db	779	KNFDALRKGRGSTSTFLNIVIELKCCNHAALIRPSEFELMGLQDEALQTLKGSGL	838
Qy	598	QLLDKMMVKLKEQGRVLIYTFQFHMLDLLEDYCTHKKWQYERIDGKVGAERQIRDRF	657
Db	839	VLLDKLLCRLKETGHRVLIIFSQVMRMLDVLDADYLRKHFPPQRLDGSIKGEMRRQALDHF	898
Qy	658	NAKNSNKFCLLSTRAGGLGINLATADTVIYDSWNPHADLQAMARHLGQTNKVMYI	717
Db	899	NAEQSDFCFLSTRAGGLGINLATADTVIIFSDWNPNQNDLQAAHRAHRIGKQNVNIY	958
Qy	718	RLINRGTIERMMQLTKKMWLEHLVVGKIKTQN-----INQEELDDI	760
Db	959	RLVTARSVEEQIVERAKQKMWLDHLVIQRMDDTGTRTLKSGNGHSSNSNPENKDDLAI	1018
Qy	761	IRYGSKELFASEDDEAGSKGIHYDDAIDKLLDRDLVEAEVSVDD--EENGFLKAFKV	819
Db	1019	LKFGAEELFKDEQ-----HDDDLVCD--IDEILRRAETRNEDPEMPADLLSAFKV	1068
Qy	820	ANFEYIDENEAALAEQRAAEKSSAGNSDRASYWHEELKDKF-----ELHQAEEELNAL	874



Db 1069 ASIAAFEEE-----PSDSVSKQDQNAAGEEDDSKDWDIIPEGFRKAIDQERAKEMEDL 1123

QY 875 -----GKRKRSRKQLVSIIEEDDLAILEDVSSDGDSEYEAESTDGEAAGQGV 920

Db 1124 YLPPRRKTAANQNEGRKAGKGKQKQADD-----SGGSDYELGS-----DGS 1168

QY 921 QTGRRPYRRKGRDNLEPTPLMEGEGRSFVLGFNQSORAIFVQTLRMRYGAGNFDWKEFV- 979

Db 1169 GDDGRP-RKGRPTMKE-----KITFTDAELRRFIRSYKKFPAPLHRMEAIAC 1216

QY 980 -PRLKQKTFEEINEYG-----ILFLKHIAEEIDENSPTFSFGVPKE-----GLRIE 1024

Db 1217 DAELQEKPLAELKRLGEMLHDCVQFLHEHKEEESKTAATDETPGAKQRRARATFSVKLG 1276

QY 1025 DVLVRIALLILVQEKVKFVEDHPGKPVFPSPRILERFPGLRSGKI-----WKEEHDKI 1076

Db 1277 GVSFNAKLLACEQELQ-----PLNEIMPSMPEERQQWSFNIKTRAPVFDVDMGIEEDTK 1331

QY 1077 MIRAVLKHGYGRWQAIVDKDELGIQELICKELNFPHISLSAAEQAGLQGGSGGSPGA 1136

Db 1332 LLCGIYQYGIGSWEQMLDPTLKLTDKILL-----1361

QY 1137 QTNQNPGSVITGNNNASADGAQVNSMFYYRDMQRRLVEFVKRVLLE-----1184

Db 1362 -----NDTRKPAKQLQTRAEY-----LLKIKNVELTKGGORRRPRAS 1403

QY 1185 KAMNVEYAEYYGLGSSSIPTEEPEAEPKIADTVGSFIEVDDEMLD---GLPKTDPIIT 1241

Db 1404 RANDAKAASQ---SASSTIDAKPHDGEDAAGDARTVA--ESSNSQVDPSPNASPHNAPAT 1457

QY 1242 SEEIMGAADVNNQARVEIAQHYNQMCKLLDENARESQAAYVNNQPPSTKV-----N 1292

Db 1458 EQH-----GDPAKKAKKSARSKKTSAASNNGNKPMMHFTANNEPRALEVLGDLDPISIFN 1511

QY 1293 ESFRALKSINGNINTI-----LSITSQSKSHEDD 1322

Db 1512 ECKEKMRPVKKALKALDQDPVSLSDQDQLQHTRD 1545

RESULT 11

T38720

chromodomain helicase hrp3 - fission yeast (Schizosaccharomyces pombe)

C;Species: Schizosaccharomyces pombe

C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Dec-2002

C;Accession: T38720

R;Gentles, S.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V. submitted to the EMBL Data Library, September 1997

A;Reference number: Z21797

A;Accession: T38720

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-1388 <GEN>

A;Cross-references: EMBL:Z99167; PIDN:CAB16277.1; GSPDB:GN00066; SPDB:SPAC3G6.01

A;Experimental source: strain 972h-; cosmid c3G6

C;Genetics:

A;Gene: SPDB:SPAC3G6.01

A;Map position: 1

A;Introns: 64/3

C;Superfamily: chromodomain helicase CHD1; chromobox homology

Query Match 20.5%; Score 1473; DB 2; Length 1388;

Best Local Similarity 29.3%; Pred. No. 9.7e-67;

Matches 429; Conservative 215; Mismatches 437; Indels 382; Gaps 42;

QY 12 SDRKPVYNLDD-----SDDDDF--VPKKDRTFEQVEAIVRTDAKENACQACGES 58

Db 99 SRHRTIRDLDEAESVTSESESDSSYGGTTPKKRSQKKSNTYVQDEIRFSSRSKGVN 158

QY 59 TNLVSCNTCTYAFHAKCLVPLPKDASVENW-----RCPECVSPLNE-----IDKILD 105

Db 159 YN-----EDAYFESFEEEEEMYEYATEVSEEPEDTRAIDVVLD 198

QY 106 CEMRPTKSSEQSSDAEPKPIFVKQYLVKWKGLSYLHCSWVPEKEFQKAYKSNHRLKTRV 165

Db 199 HRLIEGHGSGTPSEDYE-----FLIKWVNFSLHCTWEPYNNI-----SMIRGSKV 245

QY 166 NNFHRQM-----ESFNSEDDFVAI-----RPEWTTVDRIACR-EEDGELE 206

Db 246 DNHIKQVILLDREIREDPTTTREDIEAMDIEKREKRENYEYKQVDRIVAKHLNSDGSVE 305

QY 207 YLVKYKELSYDECYWESESDISTF-QNEIQRFKD-VNSRTRRSKVDVHKRNPRDFQQFDH 264

Db 306 YLVKWKQLLYDFCTWEASSIIIEPIAAATEIQAQEREESALSPSRGTNYGNSRPKYRKLEQ 365

QY 265 TPEFLK-GLLHPYQLEGLNFLRFSWSKQTHVILADEMGLKTIQSIALLA-----SLFEEN 319

Db 366 QPSYITGGELRDFQLTGVNWMAYLWHKNENGILADEMGLKTVQTVAFLSYLAHSLRQH 425

QY 320 LIPHLVIAPLSTLRNWEREFATWAPQMVVMVYFGTAQARAVIREHEFYLSKDQKKIKKK 379

Db 426 --PFLVVVPLSTVPWQETLALWASDMNCISYLGNTTSRQVIRDYEFYVDGTQK-----477

QY 380 SGQISSESQKRIKFDVLLTSYEMINLDSAVLKPIKWECEMIVDEGHRLKNKDSKLFSSLT 439

Db 478 -----IKFNLLLTYYEVLKDRSVLSNIKWQYMAIDEAHLKNSESSLYEALS 525

QY 440 QYSSNHRILLTGTPLONNDELFLMLHFDLADGKFGSLEEFQEEFKDINQEEQISRLHKML 499

Db 526 QFKNSNRLITGTPLQNNIRELAALVDFLMPGKFEIREEINLEAPDEEQEAYIRSLQEHL 585

QY 500 APHLRRVKDVMKDMPPKELILRVDSLSSLOKEYYKAIFTRNYYQLTK--KGGAQISLN 557

Db 586 QPYILRRLKKDVEKSLPSKSERILRVVELSDLQMYWYKNILTRNRYVLTSQISSGSQISLL 645

QY 558 NIMMELRKVCCHPYMLEGVE----PVIHD----ANEAFKQLLESCGKLQLLDKMMVKLKEQ 610

Db 646 NIVVELKKASNHPYLPFDGVEESWMQKINSQGRRDEVLKLJMNSGKMLLDKLLSRLRRD 705

QY 611 GHRVLIYTFQFQHMLDLLEDYCTHKKWQYERIDGKVGGAERQIRIDRFNAKNSNKKFCFLS 670

Db 706 GHRVLIQFQMVVRMLDILGDYLSLRGYPHQRLDGTVPAAVRRTSIDHFNAPNSPDFVFLLS 765

QY 671 TRAGGLGINLATADTVIYYDSDNPHADLQAMARAHRLGQTNKVMIVRLINRGTIERMM 730

Db 766 TRAGGLGINLMTADTVIIFDSDNWPNQADLQAMARAHRIQKQNHVMVYRLLSKDTIEEDVL 825

QY 731 QLTKKKMVLEHLVVG-----KLKTONINQEELDDIIRYGSKELFASEDDEAGSKGI 782

Db 826 ERARRKMILEYAIISLGVTDKQKNSKNDKFSABELSAILKFGASNMFKAENNQ-----878

QY 783 HYDDAAIDKLDRDLVEAEESVVDDEENG-----FLKAFKVANFE-----823

Db 879 -----KKLEDMNLDEILEHAEDHDTSDNVGGASMGEEFLKQFEVTDYKADVSWDDII 931

QY 824 -----YIDENEAALAEQ-----RVAAESKSSAGNSDRASYWEELL 859

Db 932 PLTEREKFEEDRLREEEALKQEIELSRRGNRPYPSSAVESPSYSGTSERKSK-KQML 990

QY 860 KDKFELHQAEEELNALGKRKRSRKQLVSIIE--DDLAGLEDVSSDGDSEYEAESTDGEAAG 917

Db 991 KDEVILLE--KEIRLL---YRAMIRYGSLEHRYNDIVKYADLTQDAHVIVKXIAADLVITAS 1045

QY 918 QGVQTGRRPYPYRRKGRDNLEPTPLMEGEGRSFVLGFNQSORAIFVQTLRMRYGAGNFDWKE 977

Db 1046 RKAVSAAEKDLSNDQSN-----NKSSRKALLITFK--GVKNINAET 1084

QY 978 FVPRLKQKTFEEINEYGILFLKHIAEEIDENSPTFSFGVPKEGLRIEDVLVRIALLILVQ 1037

Db 1085 LVQRL-----NDLDILY-----DAMPTSGYSNFQIPMHV-----1113

QY 1038 EKVKFVEDHPGKPVFPSPRILERFPGLRS-----GKIWKEEHDKIMIRAVLKHGYGRWQAI 1093

Db 1114 -----RSVHGWSCQWGPREDSMLLSGICKHGFQAWLEIR 1147

QY 1094 DDKELGIQELICKELNFPHISLSAAEQAGLQGGSGGSPGAQTNQNPQSVITGNNNAS 1153

Db 1148 DPPELKMCKIFLE-----DTKQTDNSVPKDKENKE 1178

Qy 1154 ADGAQVNSMFYRDMQRRLLVEFVKRVLLEBKAMNYEAYEEYGLGGSSSIPTEEPEAP 1213

Db 1179 ---KKVPS-----AVHLVRRGEYLLSALREH---HQNFGIKSSPAISTN----- 1216

Qy 1214 KIADTVGSFIEVDDEMLDGLPKTDPIITSEEIMGAADVNNQARVEIAQHYNQMCKLLDEN 1273

Db 1217 -----GKTQP-----KKQTAN 1227

Qy 1274 ARESVQAYVNN-----QPPSTKVNESFRALKSINGNINT-----ILSITSDQSKSHE 1320

Db 1228 RRQSGKPNVKSQAQKIESATRTSPPAISESRKKPSSKDTKIETPSREQRSQTASPVKSEK 1287

Qy 1321 DDTKPDLLNNVEMKDTAEETKPLR 1343

Db 1288 DDGNVSLNAEQKARCKELMYPVR 1310

RESULT 12

S30818

hypothetical protein YER164w - yeast (Saccharomyces cerevisiae)

C;Species: Saccharomyces cerevisiae

C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Dec-2002

C;Accession: S30818; S50667

R;Mulligan, J.T.; Dietrich, F.S.; Hennessey, K.M.; Sehl, P.; Komp, C.; Wei, Y.; Taylor, submitted to the EMBL Data Library, February 1993

A;Reference number: S30812

A;Accession: S30818

A;Molecule type: DNA

A;Residues: 1-1468 <MUL>

A;Cross-references: GB:U18917; EMBL:L10718; NID:G603377; PIDN:AAB64691.1; PID:G603404

R;Dietrich, F.S.

submitted to the EMBL Data Library, December 1994

A;Description: The sequence of S. cerevisiae cosmid 8229, 9115, 9132, 9981, and lambda

A;Reference number: S50667

A;Accession: S50667

A;Molecule type: DNA

A;Residues: 1-1468 <DIE>

A;Cross-references: EMBL:U18917; NID:G603377; PIDN:AAB64691.1; PID:G603404; GSPDB:GN0000

C;Genetics:

A;Gene: SGD:CHD1; MIPS:YER164w

A;Cross-references: SGD:S0000966; MIPS:YER164w

A;Map position: 5R

C;Superfamily: chromodomain helicase CHD1; chromobox homology

C;Keywords: DNA binding; nucleus

F;195-233/Domain: chromobox homology <CB1>

F;285-327/Domain: chromobox homology <CB2>

Query Match 20.3%; Score 1461.5; DB 1; Length 1468;

Best Local Similarity 32.1%; Pred. No. 4e-66;

Matches 410; Conservative 217; Mismatches 403; Indels 249; Gaps 46;

Qy 7 RLRISSDRKPVVNLDDSDDDDFVPKKDRTFEQVEAIVRTDAKENACQACGESTNLVSCNT 66

Db 126 RFSNRQNKTVNINIDYSDDDLLESEDD--YGSEAL----SEENVHEA-----SANP 171

Qy 67 CTYAFHAKCLV--PPLKDASVENWRCPECVSPLNEIDKILDCEMRPTKSSEQGSSDAEPK 124

Db 172 QPEDFHGIDIVINHRLKTSLEEGKVLEKTVPDLN-----NCKEN----- 210

Qy 125 PIFVKQYLWKVKGISYLVHCSWVPEKEFKAYKSNHRLK--TRVNNFHRQ-----ME 173

Db 211 -----YEFLIKWTDESHLNTW-----ETYESIGQVRLKRLDNYCKQFIIEDQQVRLD 259

Qy 174 SFNNSED----DFVAIR-----PEWTTVDRIACRE---EDG--ELEYLVKYKELSYDEC 219

Db 260 PYVTAEDIEIMDMERERRLDEFEFFHVPERIIDSQRASLEDGTSQLQYLVKWRRRLNYDEA 319

Qy 220 YWESSEDISTFQNE-IQRFKD-VNSRTRRSKDVHDHKNRPDRFQQFDHTPEFLK-GLLHPY 276

Db 320 TWENATDIVKLAPEQVQKHFNQRENKSKILPOYSNNYTSQRPREFKLSVQPPFIKGGLRDF 379

RESULT 13

T50107

chromodomain helicase hrp1 - fission yeast (Schizosaccharomyces pombe)

C;Species: Schizosaccharomyces pombe

Qy 277 QLEGLNFLRFSWSKQTHVILADEMGLGKTIQSIALLASLF--EENLIPHLVIAPLSTLRN 334

Db 380 QLTGINWMAFLWSKGDNGILADEMGLGKTVQTVAFISWLIFARRQNGPHIIVVPLSTMPA 439

Qy 335 WEREFATWAPQMNVVMYFGTAQARAVIREHEFYLSKDQKKIKKKKSGQISSESQKRIKF 394

Db 440 WLDITFEKWAPDLNCICYMGNQKSRDTIREYEFYTNPRAK-----GKKTMKF 485

Qy 395 DVLLTSYEMINLDSAVLKPIKWECMIVDEGHRLKNKDSKLFSSLTQYSSNHRILLTGTP 454

Db 486 NVLLTTYEYILKDRAELGSIKWQFMAVDEAHLKNAESSLYESNSFKVANRMLITGTP 545

Qy 455 QNNLDELFMLMHFLDAGKFGSLEEFQEEFKDINQEEQISRLHKMLAPHLRLRRVKDVMKD 514

Db 546 QNNIKELAALVNFLMPGRFTIDQEIFENQDEEQEYIHDLHRRIQPFILRRLKKDVEKS 605

Qy 515 MPPKKEILIRVDLSSLOKEYYKAIPTRNYYQVLT--KKGGAQISLNNIMMELRKVCCHPYM 572

Db 606 LPSKTERILRVELSDVQTEYYKNIILTKNYSALTAGAKG-HFSLNIMNELKKASNHPLY 664

Qy 573 LEGVEPVI-----HDANAEFKQLLESCGKLQLLDKMMVKLBQGHRLVIYTFQHML 624

Db 665 FDNAEERVLOKFGDGKMTRENVLRGLIMSSGKMVLDDQLLTRLKDGHRVLIFSQVMRML 724

Qy 625 DLLEDYCTHKKWQYERIDGKVGGAEQIRIDRFNAKNSNKFCFLLSTRAGGLGINLATAD 684

Db 725 DILGDYLSIKGINFORLDGTVPQAORRISIDHFNSPDSNDFVLLSTRAGGLGINLMTAD 784

Qy 685 TVIYDSDMNPHADLOAMARAHRLGQTNKVMYIRLINRGTIEERNMQLTKKMVLEHLVV 744

Db 785 TVVIFSDMNPQADLOAMARAHRIQGNHVMYVRLVSKDTVEEVLERARKKMILEYAI 844

Qy 745 G-----KLKTONINQEEELDDIIRYGSKELFASEDDEAGSKIHYYDDA---AIDKLL 793

Db 845 SLGVTGDKVYTKKNEPNAGELSAILKFGAGNMFTATDNQK-KLEDNLDDVLNHAEDHVT 903

Qy 794 DRDLVEABEVSVDDEEENGFLKAFKVANF-----EYIDENEAALAEQAQVAAESKSSA 846

Db 904 TPDLGESH---LGEE---FLKQFEVTDYKADIDWDDDIIPPEELKKIQ----- 945

Qy 847 GNSDRASYWHEELLKDKFELHQAEEELNALGKRKRSRKQLV-SIEEDDLAGLEDVSSDGDES 905

Db 946 -----DEEQXRKDEEYVKEQLEMMNRRDNALKKIKNSVNGDGT-----ANSDSD- 991

Qy 906 YEABSTDGEAAGQGVQTGRRPYRRKGRDNLEPTPLMEGGRSF--RVLGFNQ----- 955

Db 992 ---DSTS-----RSSRRARAN-DMDSIGESEVRALYKAILKFGNLKEILDEL 1035

Qy 956 -SQRAIFVQTLRMRYGAGNFDWKEFVPRLKQKTFFEEINEYGILFLKHIAEEIDENSPTF-- 1012

Db 1036 IADGTLPVKSFEKYGE---TYDENMEAAKDCVHEEKN-----RKEILEKLEKHATAYRA 1087

Qy 1013 ---SDGVPKEGLRIEDVLRIALLILVQEKVKFVEDHPG-KPVFPSPRILERFPGLRSGK- 1067

Db 1088 KLSGSEIKAEENQPKDNPLTRLSL--KKREKKAVLFNFKGVKSLNAESLLSRVEDLKYLKN 1145

Qy 1068 -----IWKEEHDKIMIRAVLKHGYGRWQAIVDDKELG 1099

Db 1146 LINSNYKDDPLKFSLGNWTPKPVQNWSSNWTKEDEKLLIGVFKYGYGSWTQIRDDPFLG 1205

Qy 1100 I-QLICKELNFP--HISLSAAEQAGLQGGQNGSG--GSNPGAQTNQNPGSV----- 1145

Db 1206 ITDKIFLNEVHNVPVAKKSASSSDTTPPSKKKGKITGS-----SKKVPGAHLGRRVDYL 1260

Qy 1146 ---ITGNNNASADGAQVNS 1161

Db 1261 LSFLRGLLNTKSPSADIGS 1279





QY 83 ASVENWRCPECVSPLNEIDKILDCEMRPTKSSEQSSDAEPKPIFVKQYLVKWKGLSYLH 142  
Db 193 ---EEEEVEEQVE--EVEYPIIDFVLNHRKGRAD--AQDDGPKSSY--QYLIKQWQEVSHLH 243  
QY 143 CSWVPEKEFQ--KAYKSNHRLKTRVNNFHRQ-----MESFNNSDEDFAIR----- 186  
Db 244 NTWEDYSTLSSVRGYK-----KVDNYIKQNIYDREIREDPPTTFEDIEALDIERERK 296  
QY 187 ----PEWTTVDRIILACR--EDGELEYLVKYKELSYDECYWESESDI--STFQNEIQRF-K 238  
Db 297 NMLFEYKIVERIVASETNEEGKTEYFVKWQLPYDNCTWE--DSDVIYMAPNEVYQFLO 355  
QY 239 DVNSRTRRSKDVHKKRNPDPFQQFDHTPEFLK-GLLHPYQLEGFLNFRFSWSKQTHVILA 297  
Db 356 RENSPYLPYKGVFNTRP-PYRKLEKQPSYMKGGEIRDQFLTGINWMAYLWHRNENGILA 414  
QY 298 DEMGLGKTIQSIALLASLFE--ENLIPHVLVAPLSTLRNWEREFATWAPQMNVMVYFGTA 355  
Db 415 DEMGLGKTQVTCVFLSYLVHSLKHQHPFLIVVPLSTVPAWQETLANWTPDLNSICYTGT 474  
QY 356 QARAVIREHEFYLSKDQKKIKKKSGQISSBSKQRIKFDVLLTSYBEMINLDSAVLKPIK 415  
Db 475 ESRAIL-----ESMNSRKLFKNILLTTYEYLKDKQELNNIR 511  
QY 416 WECMIVDEGHRLLKNKDSKLFSSLTQYSSNHRILLTGTPLONNLDELMLMHFLDAGKFGS 475  
Db 512 WQYLAIDEAHRLLKNSESSLYETLSQFRTANRLITGTPLQNNLKELASLVNFLMPGKFYI 571  
QY 476 LEEFQEEFKDINOEEQISRLHKMLAPHLRRVKQDVMDKMPKPKELILRVDLSSLQKEYY 535  
Db 572 RDELNFDQPNAEQERDIRDQERLQFPFILRLKKDVEKSLPSKSERILRVELSDMQTEWY 631  
QY 536 KAIFTRNYQVLT--KKGGAQISLNNIMMELRKVCCHPYMLEGVEPV-----IHDANEAFK 588  
Db 632 KNILTKNYRALTGHTDGRGQLSLNIVVELKVKVSNHPYLPFGAAEKWMMGRKMTREDTLR 691  
QY 589 QLLESCGKLQLLDKMMVKLKEQHRVLIYTOFQHMLDLLEDYCTHKKQWYERIDGKVGGA 648  
Db 692 GIIMNSGKMWLLDKLLQRLKHDGHRVLIFSQVMRMLNILGEYMSLRGYNVQRLDGTIPAS 751  
QY 649 ERQIRIDRFNAKSNKFCFLSTRAGGLGINLATADTVIYDSDWNHADLOAMARHRL 708  
Db 752 VRRVSIDHFNAPDSPDFVFLSTRAGGLGINLTADTVIIFSDWNPNQADLOAMARAHRI 811  
QY 709 QOTNKVMYRLINRGTIERMMQLTKKMVLEHLVVG-----KLKTQINQEEELDDI 760  
Db 812 GQKNHVNVYRFLSKDTVEEDILERARRKMILEYAIISLGVTEKSKNSKNDKYDAQELSAI 871  
QY 761 IRYGSKELFASDEDEAGSKGIHYDDAAIDKLLDRDLVEAEV---SVDDEENGFLKAF 817  
Db 872 LKFGASNMFKATENQK-KLENMNLDD--ILSHAEDRD--SSNDVGGGASMGGEE---FLKQF 924  
QY 818 KVANFEYID-----ENEAALEAQRVAAESKSSAGNSDRASYWEELLKDK 862  
Db 925 EVTDYKAEDLNWDDIIPEEEMERIEEEERMLAAQRAKEERERR-----EEEEEREN 975  
QY 863 FELHQA----EELNALGKRKRSLKQLVSIIEDDL-----GLEDVSSDG-DESYEAES 910  
Db 976 DEDHPSRTYKRTKTSITKRQORREEMVREKEIRLLYRAMIKFGLVDERFDTIVKEAELQA 1035  
QY 911 TDGEAAGQGVQTCRRPYRRKGRDNLPTPLMEGEGRSFRVLGFNSQRAIFVQTLMRYYA 970  
Db 1036 TDPK-----RIYLSADMVKACDEAVERLGA 1061  
QY 971 GNFDWKEFVPRLKQKTFEEINEYGILFLKHAEEIDENSPTFSQVPKEGLRIEDVLVRI 1030  
Db 1062 D--DTGNKQPR-----KAILIEF-----KGV--KNINAETVTLRV 1092  
QY 1031 ALLILVQEKVKFVEDHPGKPV--FPSRILRFPGLRSGKIWKEEHDKIMIRAVLKHGYGR 1088  
Db 1093 KDLTHLHRAVKGLD--PLKQIIGYPIRSVHSW-----NCSWGIKEDSMMLAGINKHGFGC 1145  
QY 1089 WQAIVDDKELGIQELIC-----KELNF-----PHISLSAAEQ 1120

Db 1146 WQAIKNDDPDLGLHDKIFLDEAKNDKESRYVPSAVHLVRRGEYLLSVVREHPDLFVVKTDQ 1205  
QY 1121 AGLOGQNGSGSNPGAQTNQNPQSGVITGNNNNASADGAQVNSMFYRDMQRRLLVEFVKRV 1180  
Db 1206 -----PTKRKYNRKAPTCKSSTRQTTLDGSIISNTKKSRTKKKK----- 1243  
QY 1181 LLEKAMNYEYAEYYGLGGSSSIPTEEPEAEKPIADTVGVSFIEVDDEMLDGLPKTDP 1240  
Db 1244 -----EETNRGDETS-----PEG-----TVGED--EVEEPRQAEPPKRAL 1278  
QY 1241 TSEEIMGAAVDNNQARVEIAQHYNQCKLLDENARESQAAYVNNQPPSTKVNESFRALKS 1300  
Db 1279 RSNS--GKAASNKRTTRNSMKTHSAMDTL---TAAALDAELDNM-SNEKAKEEVDHVKS 1332  
QY 1301 ING-NINTILSITSDQSKSHEDDTKPDNLNNVEMKDTAEETKPL 1342  
Db 1333 ENGESVNE--PNTEDLSLETEENT-----TVSDISPL 1362  
RESULT 15  
T33152  
hypothetical protein T04D1.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T33152  
R:Davidson, S.; Wohldmann, P.  
submitted to the EMBL Data Library, May 1998  
A:Description: The sequence of C. elegans cosmid T04D1.  
A:Reference number: 221292  
A:Accession: T33152  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2957 <DAV>  
A:Cross-references: EMBL:AF067617; PIDN:AAC17559.1; GSPDB:GN00019; CESP:T04D1.4  
A:Experimental source: strain Bristol N2; clone T04D1  
C:Genetics:  
A:Gene: CESP:T04D1.4  
A:Map position: 1  
A:Introns: 122/3; 293/3; 515/3; 1205/2; 1577/1; 2221/3; 2776/1; 2864/3  
Query Match 18.5%; Score 1332.5; DB 2; Length 2957;  
Best Local Similarity 29.3%; Pred. No. 3.8e-59;  
Matches 368; Conservative 219; Mismatches 386; Indels 281; Gaps 43;  
QY 11 RSDRKPVY-----NLDDSDDDDFVP-----KKDRTFEQVEAIVRTDA 47  
Db 895 KSYRPAVVVAKSRKKKSVDDSDDDDFYPQGRGKKKGGGKKKADPTDEAGVEGAKSDD 954  
QY 48 KENACQACGESTNLVSCNTCTYAFHAKCLVP-PLKDASVENWRCPCEVCVSPNLNIDKILDC 106  
Db 955 EDD-----EFLMKIDTPAPDPNAMV-----VEKILNM 981  
QY 107 EM-----RPTKSSEQSS-----DAEPKP 125  
Db 982 RMGKVEKKVPETEAEGSEENGEEKSEKVKGKPKENGESVGNGESKSKTSETNE 1041  
QY 126 I---FVKQYLVKWKGLSYLHCSWVPEKEFQKAYKSNHRLKTRVNNFH-RQMESFNNSDD 181  
Db 1042 IEEEEVEQFLIKWGRSYVHCEWKTAAELLEIDK---RVEAKIKRFKVKWKMSYIEDDED 1098  
QY 182 FVAIRPEWTTVDRIIL-ACREEDGELEYLVKYKELSYDECYWESESDISTFQNEIQRFKDV 240  
Db 1099 F---NSDFVIVDRVVDLITEDDQGEFVLIKWKSGLGYEEVTWEPIEMIPADKVELWRERQV 1155  
QY 241 NSRTR-RSKDQVDHKNRNPDPFQQFDHTPEFLKG-LLHPYQLEGFLNFRFSWSKQTHVILAD 298  
Db 1156 IDPAKIREK---QRPEPEEWKMKMSTSKVWKNGNLSREYQPEGVDWLLYCYNAQNCLILAD 1212  
QY 299 EMGLKTIQSIALLASLFEENLI-PHLVIAPLSTLRNWEREFATWAPQMNVMVYFGTAQA 357  
Db 1213 EMGLGKTVQTITFLSRIYDYGHPFLVVPVPLSTIQNVVREFETWT-DMNAIVYHGSAYA 1271



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 14, 2004, 01:51:41 ; Search time 31 Seconds  
(without alignments)  
2324.680 Million cell updates/sec

Title: US-10-049-137-2

Perfect score: 7187

Sequence: 1 MSSLVERLIRSDRKPVYNL.....VDVKMEEAKEEKPKNMVVD 1384

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2091	29.1	1912	1	CHD4_HUMAN	Q14839 homo sapien
2	2083	29.0	1944	1	CHD3_HUMAN	Q12873 homo sapien
3	2045.5	28.5	1982	1	CHDM_DROME	Q97159 drosophila
4	2032	28.3	1787	1	CHD3_CAEEL	Q22516 caenorhabdi
5	1758	24.5	892	1	CHD3_DROME	Q16102 drosophila
6	1621.5	22.6	1739	1	CHD2_HUMAN	Q14647 homo sapien
7	1572.5	21.9	1711	1	CHD1_MOUSE	P40201 mus musculu
8	1562.5	21.7	1709	1	CHD1_HUMAN	Q14646 homo sapien
9	1494	20.8	2004	1	CHDB_HUMAN	Q9hck8 homo sapien
10	1492.5	20.8	1967	1	CHDA_HUMAN	Q9p2d1 homo sapien
11	1473	20.5	1388	1	HRP3_SCHPO	O14139 schizosacch
12	1461.5	20.3	1468	1	CHD1_YEAST	P32657 saccharomyc
13	1422.5	19.8	2713	1	CHD5_HUMAN	Q8td26 homo sapien
14	1402	19.5	1373	1	HRP1_SCHPO	Q9us25 schizosacch
15	1139	15.8	976	1	SN21_HUMAN	P28370 homo sapien
16	1135.5	15.8	1129	1	YB95_YEAST	P38144 saccharomyc
17	1131	15.7	1027	1	ISWI_DROME	Q24368 drosophila
18	1126.5	15.7	1009	1	SN2L_CAEEL	P41877 caenorhabdi
19	1120	15.6	1359	1	STH1_YEAST	P32597 saccharomyc
20	1095.5	15.2	1703	1	SNF2_YEAST	P22082 saccharomyc
21	1040.5	14.5	1638	1	BRM_DROME	P25439 drosophila
22	1027	14.3	1647	1	SN24_HUMAN	P51532 homo sapien
23	1024	14.2	1586	1	SN22_HUMAN	P51531 homo sapien
24	922	12.8	778	1	YFK8_YEAST	P43610 saccharomyc
25	879	12.2	1514	1	YD34_YEAST	Q05471 saccharomyc
26	859.5	12.0	1489	1	YGP0_YEAST	P53115 saccharomyc
27	795	11.1	1021	1	SRD1_MOUSE	Q04692 mus musculu
28	789	11.0	1026	1	SRD1_HUMAN	Q9h4l7 homo sapien
29	784.5	10.9	1085	1	RA26_YEAST	P40352 saccharomyc
30	779.5	10.8	1867	1	MOT1_YEAST	P32333 saccharomyc
31	750.5	10.4	1131	1	YAB9_YEAST	P31380 saccharomyc
32	722	10.0	1493	1	ERC6_HUMAN	Q03468 homo sapien
33	656	9.1	1849	1	T172_HUMAN	O14981 homo sapien

RESULT 1						
CHD4_HUMAN						
ID	CHD4_HUMAN	STANDARD;	PRT;	1912	AA.	
AC	Q14839;					
DT	15-JUL-1999	(Rel. 38, Created)				
DT	15-JUL-1999	(Rel. 38, Last sequence update)				
DT	28-FEB-2003	(Rel. 41, Last annotation update)				
DE	Chromodomain helicase-DNA-binding protein 4 (CHD-4) (Mi-2 autoantigen					
DE	218 kDa protein) (Mi2-beta).					
GN	CHD4.					
OS	Homo sapiens (Human).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
OX	NCBI_TaxID=9606;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=96017437; PubMed=7575689;					
RA	Seelig H.P., Moosbrugger I., Ehrfeld H., Fink T., Renz M., Genth E.;					
RT	"The major dermatomyositis specific Mi-2 autoantigen is a presumed					
RT	helicase involved in transcriptional activation.";					
RL	Arthritis Rheum. 38:1389-1399(1995).					
CC	-!- FUNCTION: Probable transcription regulator.					
CC	-!- SUBUNIT: Central component of the nucleosome remodelling and					
CC	histone deacetylase (NURD) repressor complex.					
CC	-!- SUBCELLULAR LOCATION: Nuclear.					
CC	-!- DISEASE: One of the main antigens reacting with anti-MI-2 positive					
CC	sera of dermatomyositis.					
CC	-!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.					
CC	-!- SIMILARITY: Contains 2 chromo domains.					
CC	-!- SIMILARITY: Contains 2 PHD-type zinc fingers.					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration					
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC	the European Bioinformatics Institute. There are no restrictions on its					
CC	use by non-profit institutions as long as its content is in no way					
CC	modified and this statement is not removed. Usage by and for commercial					
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/					
CC	or send an email to license@isb-sib.ch).					
CC	-----					
CC	EMBL; X86691; CAA60384.1; -.					
DR	TRANSFAC; T05116; -.					
DR	Genew; HGNC:1919; CHD4.					
DR	MIM; 603277; -.					
DR	GO; GO:0004003; F:ATP dependent DNA helicase activity; TAS.					
DR	GO; GO:0003677; F:DNA binding; TAS.					
DR	GO; GO:0008270; F:zinc ion binding; TAS.					
DR	GO; GO:0007001; P:chromosome organization and biogenesis (sen. . .; TAS.					
DR	GO; GO:0006357; P:regulation of transcription from Pol II pro. . .; TAS.					
DR	InterPro; IPR000953; Chromo.					
DR	InterPro; IPR001410; DEAD.					
DR	InterPro; IPR002464; DEAH_box.					
DR	InterPro; IPR001650; Helicase_C.					
DR	InterPro; IPR000330; SNF2_N.					
DR	InterPro; IPR001965; Znf_PHD.					
DR	Pfam; PF00385; Chromo; 1.					
DR	Pfam; PF00271; helicase_C; 1.					

ALIGNMENTS



DR	Pfam; PF00628; PHD; 2.	
DR	Pfam; PF00176; SNE2 N; 1.	
DR	SMART; SM00298; CHROMO; 2.	
DR	SMART; SM00487; DEXDC; 1.	
DR	SMART; SM00490; HELICC; 1.	
DR	SMART; SM00249; PHD; 2.	
DR	PROSITE; PS00598; CHROMO 1; 2.	
DR	PROSITE; PS00013; CHROMO 2; 2.	
DR	PROSITE; PS00690; DEAH ATP HELICASE; 1.	
DR	PROSITE; PS01359; ZF PHD 1; 2.	
DR	PROSITE; PS00016; ZF PHD 2; 2.	
KW	Chromatin regulator; DNA-binding; ATP-binding; Helicase;	
KW	Nuclear protein; Repeat; Transcription regulation; Antigen;	
KW	Zinc-finger.	
FT	ZN_FING 370 417 PHD-TYPE 1.	
FT	ZN_FING 449 496 PHD-TYPE 2.	
FT	DOMAIN 494 594 CHROMO 1.	
FT	DOMAIN 622 697 CHROMO 2.	
FT	NP_BIND 751 758 ATP (POTENTIAL).	
FT	SITE 873 876 DEAH BOX.	
FT	DOMAIN 50 59 POLY-LYS.	
FT	DOMAIN 94 98 POLY-GLU.	
FT	DOMAIN 114 119 POLY-LYS.	
FT	DOMAIN 134 138 POLY-GLU.	
FT	DOMAIN 139 144 POLY-ASP.	
FT	DOMAIN 227 235 POLY-ALA.	
FT	DOMAIN 248 252 POLY-PRO.	
FT	DOMAIN 350 354 POLY-LYS.	
FT	DOMAIN 1052 1055 POLY-LEU.	
FT	DOMAIN 1294 1301 POLY-GLU.	
FT	DOMAIN 1665 1668 POLY-GLU.	
SQ	SEQUENCE 1912 AA; 217989 MW; 80A73300E80166CE CRC64;	
Query Match 29.1%; Score 2091; DB 1; Length 1912;		
Best Local Similarity 35.5%; Pred. No. 2.6e-98;		
Matches 520; Conservative 205; Mismatches 424; Indels 314; Gaps 44;		
QY	21 DSDDDDFVPKKDRTFEQVEAIVRTDAKENACQACGESTNLVSCNTCTYAFHAKCLVPPL 80	
Db	426 DNSEGEIILEVGGDLEED-----DHMEFCRVCKDGGELCCDTCPSSYHIHCLNPPL 480	
QY	81 KDASVENWRCBECVSP--NEIDKILDC--MRPTKSSEQSSDAE-----PKPIF---V 128	
Db	481 PEIPNGEWLCPRCTCPALKGVQKILINWKWGQPPSPTPVPRPPDADPTSPKPLEGRPE 540	
QY	129 KOYLKWKGLSVLHCSWVPEKE-----FOKAYKSNHRLKTRVNNFHRQME----- 173	
Db	541 RQFVKWQGMYSWHCSWVSELQLELHCQVFNFRNYQRKNDMDPEPPSGDFGGDEKSRKRN 600	
QY	174 ---SFNNSEDDF--VAIRPEWTTVDRI--ACREEDGELEYLVKYKELSYDECYWESE--- 224	
Db	601 KDPKFAEMEERFYRYGKPEWMMIHRILNHSVDKKGHVHYLIKWRDLPYDQASWESEDE 660	
QY	225 -SDISTFQNEIQRFKDV-----NSRTRRSKDVDHKNRPR-----DFQQFDHTPEFLK 270	
Db	661 IQDYDLFKQSYWNHRELMRGEGRPGKKLKKVKLRKLERPPETPTVDPTVKYERQPEYLD 720	
QY	271 ---GLLHPYQLEGLNFLRFSWSKQTHVILADEMGLGKTIQSIALLASLFEE--NLIPHLV 325	
Db	721 ATGGTLHPYQMEGLNWLRFWSAQGTDTILADEMGLGKTVQTAVFLSYLYKEGSHKGPFLV 780	
QY	326 IAPLSTLRNWEREFATWAPQMNVMVYFGTAQARAVIREHEFYLSKDQKIKKKKSGQISS 385	
Db	781 SAPLSTIINWEREFEMWAPDMYVVTYVGDKDSRAIIRENEFSFEDNAIRGGKKASRM--- 837	
QY	386 ESKQRIKFDVLLTSYEMINLDSAVLKPIKWECMIVDEGHRLKNKDSKLFSSLTOYSSNH 445	
Db	838 -KKEASVKFHVLLTSYELITIDMAILGSIWACLIVDEAHLKKNQSKFFRVNLNGYSLOH 896	
QY	446 RILLTGTPQNLDLDFMLMHFLDAGKFGSLEEFQEEFKDINQEEQISRLHKMLAPHLLR 505	
Db	897 KLLLTGTPQNLDLDFMLNFLTPTFRFHNLEGFLEEFADIAKEDQIKKLDMLGPHMLR 956	

QY	506 RVKDVKMDMPKKELILRVDLSSLQKEYYKAIFTRNYQVLTKKGA-QISLNNIMMELR 564	
Db	957 RLKADVFKMPSKTELIIVRVELSPMQKKYKYILTRNFEALNARGGQVSLNVMWLK 1016	
QY	565 KVCCHPYMLEGVEPVIHDANEAFKQ-----LLESCGKLQLLDKMMVKLKEQGHVRL 615	
Db	1017 KCCNHPYLF---PVA--AMEAPKMPNGMYDGSALIRASGKLLLLQKMLKNLKEGGHRLV 1070	
QY	616 IYTFQHMLDLLEDYCTHKWKQYERIDGKVGGAERQIRIDRFNAKNSNKFCLLSTRAGG 675	
Db	1071 IFSQMTKMLDLEDFLEHEGYKVERIDGGITGNMRQBAIDRFNAPGAQQFCFLSTRAGG 1130	
QY	676 LGINLATADTVIIYDSWNPHADLQAMARAHRLGQTNKVMYRLINRGTIIEERMMLTKK 735	
Db	1131 LGINLATADTVIIYDSWNPHNDIQAFSRAHRIGONKKVMYRVFVTRASVEERITQVAKK 1190	
QY	736 KMWLEHLVVGK---LKTQINQBEELDDIIRYGSKELFASEDDDEAGKSGK-----IHVD 785	
Db	1191 KMWLTHLVVRPGLSGKTGSMKQELDDILKFGTEELFKDEATDGGGDNKEGEDSSSVIHVD 1250	
QY	786 DAAIDKLLDRDLVEAEVSDDEEENGFLKAFKVANFEYIDENEAALAEQAVAAESKSS 845	
Db	1251 DKAIERLLDRNQDETEDEL--QGMNEYLSSEFKVAQVYVREEEMGEEEEVEEREIIKQEEES 1308	
QY	846 AGNSDRASYWEELLKDKFELHQAELNALGKRKRKQL-----VSIEDDDLAGLEDVSSD 901	
Db	1309 VD---PDYWEKLLRHHYEQOQEDLARNLGKGRIRKQVNYNDGSEQEDRD---WQDDQSD 1361	
QY	902 GDESYEAESTDG-EAAGQGVQTRRRPYR---RKGRDNLBPTPLMEGEGRSFRVLGFNQSQ 957	
Db	1362 NQSDYSVASEEGDEDFDERSEAPRRPSPRKGRLNRDKDKPLP-PLLARVGNIEVLGFNARQ 1420	
QY	958 RAIFVQTLMRYG-----AGNFDWKEFVPRLKQKTFEEINEYGIILFKHIAEEIDENSPTF 1012	
Db	1421 RKAFNLAIMRYGMPPPQDAFTQW--LVRDLRGKSEKEFEKAYVSLFMRHLCEPGADGAETF 1478	
QY	1013 SDGVPKEGLRIEDVLVRIALLILVQEKVKFVEDHPGKVPFPSPRILERFPGLRSGKIWKEE 1072	
Db	1479 ADGVPREGLSRQHVLRIGVMSLRKKKVOEFE----- 1510	
QY	1073 HDKIMIRAVLKHGYGRWQAIVDDKELGIQELICKELNPFHISLSAAEQAGLQONGSGGS 1132	
Db	1511 -----HVNGRW-----SMP--ELAEVEENKMSQPGS--P 1536	
QY	1133 NPGAQTNQNPGSVITGNNNASADGAQVNSMFYIRDQMQRRLVEFVKRVLLEKAMNVEYA 1192	
Db	1537 SPKTPPTPSTPGD-----TQPT----- 1553	
QY	1193 EEEYGLGSSSIPTEEPEAEPIADTVGVSFIEVDDEMLDGLPKTDPITSEEIMGAAVDN 1252	
Db	1554 -----PAPVPPAE-----DGIK----- 1565	
QY	1253 NQARVEIAQHYNQMKLLDENARESVOAYVNNQPPSTKVNESFRALKSINGNI-NTILSI 1311	
Db	1566 -----IEN-----SLKEESIEGEKEVKSTAPETAIEC 1594	
QY	1312 TSDQSKSHEDDT-----KPDLLNNVEMKDTAE---ETKPLRGGVVDLNVVEGEENIAE 1360	
Db	1595 TQAPAPASEDEKVVVEPPPEGEEKVEKAEVKERTEEPMEETP--KGAADVEKVE-EKSAID 1651	
QY	1361 ASGSVDVKMEAEAKEEEKPKNMVV 1383	
Db	1652 LTPIVVEDKBEKEKEKEKEVML 1674	

RESULT 2  
CHD3\_HUMAN  
ID\_CHD3\_HUMAN STANDARD; PRT; 1944 AA.  
AC Q12873;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Chromodomain helicase-DNA-binding protein 3 (CHD-3) (Mi-2 autoantigen)

DE 240 kDa protein) (Mi2-alpha).

GN CHD3.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Fetal;

RX MEDLINE=97470991; PubMed=9326634;

RA Ge Q., Nilasena D.S., O'Brien C.A., Frank M.B., Targoff I.N.;

RT "Molecular analysis of a major antigenic region of the 240 kD protein

RT of Mi-2 autoantigen.";

RL J. Clin. Invest. 96:1730-1737(1995).

CC -!- FUNCTION: Probable transcription regulator.

CC -!- SUBUNIT: Central component of the nucleosome remodelling and

CC histone deacetylase (NuRD) repressive complex.

CC -!- SUBCELLULAR LOCATION: Nuclear.

CC -!- DISEASE: One of the main antigens reacting with anti-MI-2 positive

CC sera of dermatomyositis.

CC -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.

CC -!- SIMILARITY: Contains 2 chromo domains.

CC -!- SIMILARITY: Contains 2 PHD-type zinc fingers.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AF006515; AAB87383.1; -.

DR EMBL; U08379; AAC50228.1; -.

DR Genew; HGNC:1918; CHD3.

DR MIM; 602120; -.

DR GO; GO:0004003; F:ATP dependent DNA helicase activity; TAS.

DR GO; GO:0003677; F:DNA binding; TAS.

DR GO; GO:0008270; F:zinc ion binding; TAS.

DR GO; GO:0007001; P:chromosome organization and biogenesis (sen. . ; TAS.

DR GO; GO:0006357; P:regulation of transcription from Pol II pro. . ; TAS.

DR InterPro; IPR000953; Chromo.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002464; DEAH box.

DR InterPro; IPR001650; Helicase\_C.

DR InterPro; IPR000330; SNF2 N.

DR InterPro; IPR001965; Znf\_PHD.

DR Pfam; PF00385; chromo; 1.

DR Pfam; PF00271; helicase\_C; 1.

DR Pfam; PF00628; PHD; 2.

DR Pfam; PF00176; SNF2 N; 1.

DR SMART; SM00298; CHROMO; 2.

DR SMART; SM00487; DEXDC; 1.

DR SMART; SM00490; HELIC; 1.

DR SMART; SM00249; PHD; 2.

DR PROSITE; PS00598; CHROMO 1; 1.

DR PROSITE; PS50013; CHROMO 2; 2.

DR PROSITE; PS00690; DEAH ATP HELICASE; 1.

DR PROSITE; PS01359; ZF\_PHD 1; 2.

DR PROSITE; PS50016; ZF\_PHD 2; 2.

KW Chromatin regulator; DNA-binding; ATP-binding; Helicase;

KW Nuclear protein; Repeat; Transcription regulation; Antigen;

KW Zinc-finger.

FT ZN FING 379 426 PHD-TYPE 1.

FT ZN FING 456 503 PHD-TYPE 2.

FT DOMAIN 494 594 CHROMO 1.

FT	DOMAIN	631	673	CHROMO 2.	
FT	NP_BIND	761	768	ATP (POTENTIAL) .	
FT	SITE	883	886	DEAH BOX.	
FT	DOMAIN	206	221	POLY-ALA.	
FT	DOMAIN	243	246	POLY-PRO.	
FT	DOMAIN	355	358	POLY-LYS.	
FT	DOMAIN	434	446	POLY-GLU.	
FT	DOMAIN	697	703	POLY-LYS.	
FT	CONFLICT	121	126	GEDGG -> PHFQOK (IN REF. 2) .	
FT	CONFLICT	309	312	MISSING (IN REF. 2) .	
FT	CONFLICT	653	653	W -> G (IN REF. 2) .	
SQ	SEQUENCE	1944	AA; 220691	MW; 0712E6F1D5F7D335 CRC64;	
Query Match					
Best Local Similarity 29.0%; Score 2083; DB 1; Length 1944;					
Matches 520; Conservative 211; Mismatches 415; Indels 370; Gaps 41;					
Qy	52	CQACGESTNLVSCNTCTYAFHAKCLVPPPLKDASVENWRCPECVSP	--NEIDKILDCEMR	109	
Db	459	CRVCKDGGELLCCDACISSYHYHCLNPPLPDIPNGEWLCP	RCTCPVLKGRVQKILHWRWG	518	
Qy	110	-----PTKSSEQSSDA-EPKPI----	FVKQYLVKWKGLSYLHCSWVPEKEFQKAYKSNH	159	
Db	519	EPPVAVPAPQADGNPDVPPRPRLQGRSEREFFVKWVGLSYWHCSWAKELQLEIFHLVMY	578		
Qy	160	RLKTRVNNFHRQME-SFNNSDD-----	FVAIRPEWTTVDRI	196	
Db	579	RNYQRKNDMDPEPPPLDYGSGEDGSKDKRKVKDPHYAEMEEKYRFGIKPEWMTVHRIIN	638		
Qy	197	ACREEDGELEYLVKYKELSYDECYW-ESESDISTFQNEIQ----	RFKDVNSRTRRSKDVDH	252	
Db	639	HSVDKGNHYHLVKWRDLPYDQSTWEEDEMNIPEYEEHKQSWRHRRELIMGEDPAQPRKY	698		
Qy	253	KRNPRDFQ-----	QPDHTPEFLK----	GLLHPYQLEGLNFLRFSWSKQTHVI	295
Db	699	KKKKKELQGDGPPSSPTNDPTVKYETQPRFITATGGTLHMYQLEGLNWLRFWSAQGTDTI	758		
Qy	296	LADENGLGKTIQSIALLASLFEENLI--	PHLVIAPLSTLRNWEREFATWAPQNMVVMYFG	353	
Db	759	LADENGLGKTIQTIIVFLYSLYKEGHTKGPFLVSAPLSTIINWEREFQWAPKFYVVTYT	818		
Qy	354	TAQARAVIREHEFYLS----	KDOKKIKKKKSGQISSESKQRIKFDVLLTSEYEMINLDSA	409	
Db	819	DKDSRAIIRENEFSFEDNAIKGGKAFKMK-----	REAQVKFHVLLTSYELITIDQA	870	
Qy	410	VLKPIKWECEMIVDEGHRLLKNKDSKLFSSLTQYSSNHRILLTGTPLQNNLDELFLMHFLD	469		
Db	871	ALGSIRWACLVDDEAHLKKNQSKFFRVLNGYKIDHKLLLTGTPLQNNLEELFHLNFLT	930		
Qy	470	AGKFGSLEEFQEEFKDINQEEQISRLHKMLAPHLRLRRVKDVMKMPKPKELILRVDLSS	529		
Db	931	PERFNNLEGFLEEFADISKEDQIKKLHDLGPHMLRRLKADVFKNMPAKTELIVRVELSP	990		
Qy	530	LQKEYYKAI FTRNYQVLTKKGA-QISLNNIMMELRKVCCHPYMLEGVEPVIHDANEAFK	588		
Db	991	MQKKYKYI LTRNFEALNSRGGNQVSLNIMMDLKCCCNHPYLF----	PVA--AMESPK	1044	
Qy	589	-----QLLESCGKLQLLDKMMVKLKEQGHRLVIYTFQFHMLDLLEDYCTHKKWQYE	639		
Db	1045	LPSGAYEGGALIKSSGKMLLQKMLRKLKEQGHRLVIFSQMTKMLDLEDFLDYEGYKYE	1104		
Qy	640	RIDGKVGGAERQIRIDRFNAKNSNKFCLLSTRAGGLGINLATADTVIYDSWNPHADL	699		
Db	1105	RIDGGITGALRQEAIDRFNAPGAQQFCFLSTRAGGLGINLATADTVIIFDSWNPHNDI	1164		
Qy	700	QAMARAHRLGQTNKVMYRLINRGTEIERMMQLTKKKMVLHVLVVGK--	--LKTQINQEE	756	
Db	1165	QAFSRAHRIGQANKVMYRFVTRASVEERITQVAKRKMVTHLVVRPGLSGKAGSMKQE	1224		
Qy	757	LDDIIRYGSKELFASEDDEAGK--	--SGKIHDDAAIDKLLDRDLVEAEVSVDDDEENG	813	
Db	1225	LDDILKFGTEELFKOENEGENKEEDSSVIHYDNEAIARLLDRNQDATEDTDV--	--QNMNEY	1282	







Db 1630 EVAK-----TEPKETKOPEVKEEPK-----TEEEK 1656

QY 1318 SHEDDTKP-----DLNNVEMKDTAETKPLRGGVVDLNVVVEGENIAEASGSVDV 1367

Db 1657 EKVDKKPIPTTVIDDDDDVMIVKEDGELEKPSASSPKDQKAVAAA--TSAATGATGK 1714

QY 1368 KMEEAKEEEKPKNM 1381

Db 1715 GAEDSLEVLKRKF 1728

RESULT 4

CHD3\_CAEEL STANDARD; PRT; 1787 AA.

AC Q22516; Q18794;

DT 15-JUL-1999 (Rel. 38, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Chromodomain helicase-DNA-binding protein 3 homolog (CHD-3).

GN CHD-3 OR T14G8.1.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20530482; PubMed=11076750;

RA von Zelewsky T., Palladino F., Brunschwig K., Tobler H., Hajnal A.,

RA Mueller F.;

RT "The C. elegans Mi-2 chromatin-remodelling proteins function in vulval

RT cell fate determination.";

RL Development 127:5277-5284(2000).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RA Matthews P., McMurray A.;

RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: Chromatin-remodelling protein that function in vulval

CC cell fate determination.

CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).

CC -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.

CC -!- SIMILARITY: Contains 2 PHD-type zinc fingers.

CC -!- SIMILARITY: Contains 2 chromo domains.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AF308444; AAG29837.1; -.

DR EMBL; Z67884; CAA91810.1; -.

DR EMBL; Z67881; CAA91810.1; JOINED.

DR EMBL; Z67881; CAA91798.1; -.

DR EMBL; Z67884; CAA91798.1; JOINED.

DR PIR; T20160; T20160.

DR WormPep; T14G8.1; CE03657.

DR InterPro; IPR000953; Chromo.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002464; DEAH\_box.

DR InterPro; IPR001650; Helicase\_C.

DR InterPro; IPR000330; SNF2\_N.

DR InterPro; IPR001965; Znf\_PHD.

DR InterPro; IPR001841; Znf\_ring.

DR Pfam; PF00385; chromo; 1.

DR Pfam; PF00271; helicase\_C; 1.

DR Pfam; PF00628; PHD; 2.

DR Pfam; PF00176; SNF2\_N; 1.

DR SMART; SM00298; CHROMO; 2.

DR SMART; SM00487; DEXDC; 1.

DR SMART; SM00490; HELICC; 1.

DR SMART; SM00249; PHD; 2.

DR SMART; SM00184; RING; 2.

DR PROSITE; PS00598; CHROMO\_1; FALSE\_NEG.

DR PROSITE; PS50013; CHROMO\_2; 1.

DR PROSITE; PS00690; DEAH\_ATP\_HELICASE; 1.

DR PROSITE; PS01359; ZF\_PHD\_1; 2.

DR PROSITE; PS50016; ZF\_PHD\_2; 2.

KW Chromatin regulator; Nuclear protein; Repeat; Helicase; DNA-binding;

KW ATP-binding; Zinc-finger.

FT DOMAIN 59 62 POLY-LYS.

FT ZN\_FING 265 312 PHD-TYPE 1.

FT ZN\_FING 328 375 PHD-TYPE 2.

FT DOMAIN 373 476 CHROMO 1.

FT DOMAIN 501 583 CHROMO 2.

FT DOMAIN 1287 1291 POLY-ARG.

FT NP\_BIND 641 648 ATP (POTENTIAL).

FT SITE 763 766 DEAH\_BOX.

SQ SEQUENCE 1787 AA; 205254 MW; 1EFCE1FFECE59740 CRC64;

Query Match 28.3%; Score 2032; DB 1; Length 1787;

Best Local Similarity 33.5%; Pred. No. 2.3e-95;

Matches 511; Conservative 229; Mismatches 438; Indels 346; Gaps 45;

QY 19 NLDDSDDDDF-VPKKDRTFEQVEAIVRTDAKENA--CQACGESTNLVSCNTCTYAFHAKC 75

Db 295 NMEQPEGDWSCPCEEHGPDVLIVEEEPAKANMDYCRICKETSNILLCDTCPSSYHAYC 354

QY 76 LVPLKXASVENWRCPECV--SPLNEIDKILDC-----EMRPTKSSEQSSD----- 120

Db 355 IDPPLTEIPEGEWSCPRCIIPPEAQRIEKILSWRWEISYPEPLECKEGEASKDDVFLK 414

QY 121 --AEPKPIFVKQYLVKWKGLSYLHCSWVPEKEFQKAYKSNHRL---KTRVNN---FHRQM 172

Db 415 PPRKMEPRREREFFVKWKYLAYWQCEWLSETLMDVYFTALVRMYWRKVDSENPPIFEEST 474

QY 173 ESFNNSEDD-----FVAIRPEWTTVDRII-----ACREEDGELEYLVKYKELSYDE 218

Db 475 LSRHSHDHPYKLRERFYQYGVKPEWMQIHRIINHLSYAKSQD----YLVKWKELSYEH 530

QY 219 CYWE-SESDISTFQNEIQRF-----KDVNSRTRRSK-----DVDHKKR 254

Db 531 ATWERDDTDIANYEDAIKYWHHRERMLNDEVPRNVQMKIAQREAKGLGPKEDVTSRR 590

QY 255 NPRD-----FQQFDHTPEFLK---GLLHPYQLEGFLRFSWSKQTHVILADEMGLGKTIQ 307

Db 591 KGREKIDILKKYEVQPDFISSETGGNLHPYQLEGINWLHCHWSNGTDAILADEMGLKTVQ 650

QY 308 SIALLASLFEENLI--PHLVIAPLSTLRNWEREFATWAPQMVNVYFGTAQARAVIREHE 365

Db 651 SLTFLYTLMKEGHTKGPFLIAAPLSTIINWEREAELWCPDFYVVTYVVDRESRMVIREHE 710

QY 366 FYL-----SKDQKKIKKKKSGQISSESKQRIKFDVLLTSYEMINLDSAVLKPIKWECEMIV 421

Db 711 FSFVDGAVRGGPKVSKIKT-----LENLKFHVLLTSYECINMDKAILSSIDWAALVV 762

QY 422 DEGHRLLKNKDSKLFSSLTQYSSNHRILLTGTPQLQNNLDELFLMHFLDAGKFGSLEEFQE 481

Db 763 DEARLKNQSTFFKRLREYNIQYRVLLTGTPQLQNNLEELFHLNPLAPDRFNQLESFTA 822

QY 482 EFKDINQEEQISRLHKMLAPHLRLRVKVKDMKMPKPKELILRVDLSSLQKEYYKAIFTR 541

Db 823 EFSEISKEDQIEKLNLLGPHMLRLKADVLTMGPSKQELIVRVELSAMQKKYKYNILTR 882

QY 542 NYQVL-TKGGGAQISLNNIMMELRKVCCHPYM-----LEGVEPVTHDANEAFKQLLESCG 595

Db 883 NFDALNVKNGTQMSLINIIMELKKCCNHPYLFMKACLEA--PKLKNMGVEGSALIKNAG 940

QY 596 KLQLLDKXMMVKLKEQGHRVLIYTFQFHMLDLLEDYCTHKWKQYERIDGKVGGAERQIRID 655

Db 941 KVVLLQKMLRKLKDGGRVLIIFSQMTNMLDILEDFCDVEGYKYERIDGSIITGQQRQDAID 1000

QY 656 RFNAKNSNKFCLLSTRAGGLGINLATADTVIIYDSWNPHADLQAMARAHRLGQTNKVM 715

Db 1001 RYNAPGAKQFVLLSTRAGGLGINLATADFTVIYDSWNPNDIOAFSRAHRLGQKHVM 1060

QY 716 IYRLINRGTIIEERMMLTKKQWLEHLV-----GKLTQNTINQEEELDDIIRYGSKELFAS 771

Db 1061 IYRFVTKGSVEERTSVAKKQMLLTHLVVRAGLGAKDGKSMKSTELDDVLRWGTEELFKE 1120

QY 772 ED-----DEAGKSGK-----IHYDDAAIDKLDRDLVEAAEVSVDDEE-----NGFLK 815

Db 1121 EEAPVEGADGEGTSKKPNEQEIIVWDDAAVDFLLDRN---KEEKGQDGEKKEHWTNEYLS 1177

QY 816 AFKVANFEYIDENEAAALEAQRVAEKSAGNSDRASYWEEELKDKFELHQAEELNALG 875

Db 1178 SFKVATY---NTKEADDDADDEDETEVIEKGTTEEQDPNYWEKLLKHHYEQDQETELQKLG 1234

QY 876 KRKRSRKQVSIIEED---DLAGLEDVSDGDESYEAESTDGEAAGQGVQTRRRPYRRKGR 932

Db 1235 KGRVRRQVNYASENMGMQDWSAQNNOQEEDDGSEYGSNGELLQTDDEYERRRRREER 1294

QY 933 DNLEPTPLMEGEGRSFRVLGFGNOSQRAIFVQTLRMRYGAGNFD-----WKEFVPRLKQKTF 987

Db 1295 SEKLP-PLLAKVNGQIEVLGFNPRQRKAFYNAMVRMGMPQDLTQSSWQ--VRDLRNKSE 1351

QY 988 EEINEYGILFLKHTAEIIDENSPFSDGVPKEGLRIEDVLVRIALLILVQEKV---KF- 1042

Db 1352 KVFKAYSSLFMRHLCPEVPVNDSDSFMDDGVPREGLNRQAVLSRIGLSILRKVKQVEFEKEN 1411

QY 1043 ----- 1042

Db 1412 GEWSMPETREKMLATAAQASVSNLPGMIKIKEEPIDIDETPMDVDQSNITKTEELASEVK 1471

QY 1043 VEDHPGKVPFESRI-----LERPPGLRSGK---IWKEHDKIMIRAVLKHG 1085

Db 1472 VEEEPKAPRLPYKFNICDGGYTELHSLWINEEKVARNGKEYEIMHRRHDFWLLAAVAVYG 1531

QY 1086 YGRWQAIVDDKELGIGELICKELNFPHISLSAAEQAGLQGGNGSGSNPGAQTNQNPGSV 1145

Db 1532 YGRYQ-----INFQDI---MNDPKFSIV-----NEPFKQTGADPATN 1565

QY 1146 ITGNNNASADGAQVNSMFYRDMQRRLVEFVKRVLLEKAMNYEABEYVYGLGSSSIP 1205

Db 1566 FADVKN-----KFLARFKLLEQSL----- 1585

QY 1206 TEEPEAEPKIADTVGVSFIEVDDEMLDGLPKTDPITSEEIMGAADVNNQARV---EIAQ 1261

Db 1586 -----VIEQLRRAAHNKQOSPQDVQVQLAQ 1611

QY 1262 HYNQMCKLLDEN---ARESVOAYVNNQPSTKVNESFRALKSINGNINTILSITSDQSKS 1318

Db 1612 HFSELEHTADAHVNIARES-----NNG-----NRNANAILHKCLAQLDD 1650

QY 1319 HEDDTKPDNLNNVEMKDTAEETKPL 1342

Db 1651 LLSDLKTDV--ARLPATISQVRPV 1672

RESULT 5

CHD3 DROME STANDARD; PRT; 892 AA.

AC O16102; Q8SYJ8; Q9VVZ3;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Chromodomain helicase-DNA-binding protein 3.

GN CHD3 OR CG9594.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Berkeley;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt R.G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).

RN [2]

RP REVISIONS.

RX MEDLINE=22426069; PubMed=12537572;

RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,

RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,

RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,

RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,

RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,

RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,

RA Lewis S.E.;

RT "Annotation of the Drosophila melanogaster euchromatic genome: a

RT systematic review.";

RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=Berkeley; TISSUE=Embryo;

RX MEDLINE=22426066; PubMed=12537569;

RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,

RA George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,

RA Rubin G.M., Celniker S.E.;

RT "A Drosophila full-length cDNA resource.";

RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).

RN [4]

RP SEQUENCE OF 1-812 FROM N.A.

RX MEDLINE=97470991; PubMed=9326634;

RA Woodage T., Basrai M.A., Baxevanis A.D., Hieter P., Collins F.S.;

RT "Characterization of the CHD family of proteins.";

RL Proc. Natl. Acad. Sci. U.S.A. 94:11472-11477(1997).

CC -!- FUNCTION: Possible transcription activator (By similarity).

CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).

CC -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.

CC -!- SIMILARITY: Contains 2 chromo domains.

CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.

CC -!- CAUTION: Ref.3 sequence differs from that shown due to a stop

CC codon in position 628 and a frameshift in position 720.





DR	Genew; HGNC:1917; CHD2.	
DR	MIM; 602119; --	
DR	GO; GO:0004003; F:ATP dependent DNA helicase activity; TAS.	
DR	GO; GO:0003677; F:DNA binding; TAS.	
DR	GO; GO:0007001; P:chromosome organization and biogenesis (sen. . .; TAS.	
DR	GO; GO:0006357; P:regulation of transcription from Pol II pro. . .; TAS.	
DR	InterPro; IPR000953; Chromo.	
DR	InterPro; IPR001410; DEAD.	
DR	InterPro; IPR001650; Helicase_C.	
DR	InterPro; IPR000330; SNF2_N.	
DR	Pfam; PF00385; chromo; 2.	
DR	Pfam; PF00271; helicase C; 1.	
DR	Pfam; PF00176; SNF2 N; 1.	
DR	SMART; SM00298; CHROMO; 2.	
DR	SMART; SM00487; DEXDC; 1.	
DR	SMART; SM00490; HELICC; 1.	
DR	PROSITE; PS00598; CHROMO 1; 2.	
DR	PROSITE; PS50013; CHROMO_2; 2.	
DR	PROSITE; PS00690; DEAH_ATP_HELICASE; FALSE NEG.	
KW	Hydrolase; Helicase; Nuclear protein; ATP-binding; DNA-binding;	
KW	Repeat.	
FT	DOMAIN 1 74 SER-RICH.	
FT	DOMAIN 121 139 SER-RICH.	
FT	DOMAIN 261 353 CHROMO 1.	
FT	DOMAIN 378 456 CHROMO 2.	
FT	NP BIND 509 516 ATP (POTENTIAL).	
FT	SITE 617 620 DEAH BOX.	
SQ	SEQUENCE 1739 AA; 200560 MW; D62089C25EBAA40F CRC64;	
Query Match 22.6%; Score 1621.5; DB 1; Length 1739;		
Best Local Similarity 30.7%; Pred. No. 1.5e-74;		
Matches 453; Conservative 229; Mismatches 493; Indels 299; Gaps 43;		
Qy	11 RSDRKPVNLDSDDDDFVPKK-----DRTFEQVEAIVRTDAKENACQCG 56	
Db	199 RGRKKQDSSDEDDDDDEAPKRQTRRRRAKNVSYKEDDDFE-----TDS-DDLIEMTG 250	
Qy	57 ESTNLVSCNTCTYAFHAKCLVPPLKDASVENWRCPECVSPNLNEIDKILDCEMRPTKSSEQ 116	
Db	251 EGVDEQQDNSET-----IEKVLD--SRLGKKGAT 277	
Qy	117 GSS-----DAEPKPIFVK-----QYLVKWKGLSYLHCSWVPEKEFQ----KAYKS 157	
Db	278 GASTTVYAIFANGDPSGDFDTEKDEGEIQYLIKWKGSYIHSTWESEESLQQQKVKGK 337	
Qy	158 NHRLKTRVNNFHR-----QMESFNNSDEDDFVAIRPEWTTTVDRIACREE----- 201	
Db	338 LENFKKKEDEIKQWLGVSPEDVEYFNCQQLASELNKQYQIIVERVIAVKTSTLGTQTD 397	
Qy	202 -----DGELEYLVKYSDECYWESESDI-STFQNEIQRFKDVNSR-----TRR 246	
Db	398 FPAHSRKPAAPSNEPEYLCCKWMGLPYSECSWEDEALIGKKFQNCIDSFHSRNNSKTIPTRE 457	
Qy	247 SKQVDHKNRPDRFQQDHTPEFLKG---LLHPYQLEGLNFLRFSWSKQTHVILADEMGLG 303	
Db	458 CKAL--KORPR-FVALKKQPAYLGGENLELRDYQLEGLNWLASHWCKNNSVILADEMGLG 514	
Qy	304 KTIQSIALLASLFEENLI--PHLVIAPLSTLRNWEREFATWAPQMVVMVYFGTAQARAVI 361	
Db	515 KTIQTISFLSYLFHQHLYGPFLIVVPLSTLTSWQREFEIIWAPEINVVVYIGDLSMRNTI 574	
Qy	362 REHEFYLSKDQKKIKKKSGQISSSESKQRIKFDVLLTTSYEMINLDSAVLKPIKWECMIV 421	
Db	575 REYEWI-----HSQTKRLKFNALITTYEILLKDXTVLGSINWAF LGV 616	
Qy	422 DEGHRLKNKSKLSFSSLTQYSSNHRILLTGTPLQNNLDELFMHFLDAGKFGSLEEFQE 481	
Db	617 DEAHRLKNDDSLLYKTLIDFKSNHRLITGTPLQNSLKELWSLLHFIMPEKFEFWEDFEE 676	
Qy	482 EFKDINQEEQISRLHKMLAPHLLRRVVKDVMKMPKKEILIRVDSLSSLOKEYYKAIFTR 541	
Db	677 DHGK-GRENGYQSLHKVLEPFLLRVRVKDVEKSLPAKVEQILRVEMSALQKQYKWLTR 735	

Qy	542 NYQVLTG-KGGAQISLNNIMMELRKVCCHPYMLEGVEP-VIHDANEAFKQLLESCGKLQL 599	
Db	736 NYKALAKGTRGSTSGFLNIWMLKCCNHCYLKPPENRENGQEILLSLIRSSGKLIL 795	
Qy	600 LDKMWVKLKEQGHRLVIYTFQFHMLDLLEDYCTHKKWQYERIDGKVGGAERQIRIDRENA 659	
Db	796 LDKLLTRLRERGNRVLIFSVMVRLDILAELYTIKHYPQRLDGSIKGEIRKQALDHENA 855	
Qy	660 KNSNKFCELLSTRAGGLGINLATADTVIIYDSWNPHADLQAMARAHRLGQTNKVMYRL 719	
Db	856 DGSEDFCELLSTRAGGLGINLASADTVVIFDSWNPQNDLQAQARAHRIGQKKQVNIYRL 915	
Qy	720 INRGTIBERMMQLTKKQVLEHLVV-----GKLKTONINQEEELDDIIRYG 764	
Db	916 VTGKTVEEIIERAKKQVLDHLVIQRMDDTTGRTILENNSGRSNSNPFNKEELTAILKFG 975	
Qy	765 SKELFASEDDEAGKSGKIHYYDDAAIDKLLDRDLVEAEVSVDDDEENGFLKAFKVANFEY 824	
Db	976 AEDLFKELEGESEP-----QEMDIDEIL--RLAETRENEVSTSATDELLSQFKVANFAT 1028	
Qy	825 IDENEAALAEQAVAAESKSSAGNSDRASYWEELL-----KDKFELHQAEEELNALGK 876	
Db	1029 MEDEE--ELEER-----PHKD---WDEIIPPEQKKVVEERQKELEEIYMLPR 1072	
Qy	877 RKRSRQQLVSIIEEDDLAGLEDVSSDGDSEYAE-STDGEAAGQGVQTRRRPYRRKGRDNL 935	
Db	1073 IRSSTKK-----AQTNDSDDSTESKRQAQRSSASESETEDSDDDKKPKRR----- 1117	
Qy	936 EPTPLMEGEGRSFR---VLGFNQSQRAIFVQTLRMRYGAGNFDWKEFVPR--LKQKTFEE 989	
Db	1118 -----GRPRSVRKDLVEGTDAEIRRFKAYKKFGL-PLERLECLARDAELVDSKSVAD 1169	
Qy	990 INEYGILF-----LKHIAEEIDENSPTFSDGVPKPEGLRIEDVLVRIALLILVQEKVKF 1042	
Db	1170 LKRLGELIHNCSVSAMQYEEQLKENASEGKPGKRRGPTIKISGVQVNVKSIIOHEEF 1229	
Qy	1043 VEDHPGKPVFPSR-----ILERFPGLRSGKIWKEEHDKIMIRAVLKHGYGRWQAIVDDKE 1097	
Db	1230 EMLHKSIPVDPEEKKKYCLTRVKAHFDVEGWVEDDSRLLLGIYEHGYGNWELIKTDPE 1289	
Qy	1098 LGIQELIC-----KELNFPHISLSAAEQAGLQGGNGSGSNPGAQTNQNPGSVIT 1147	
Db	1290 LKLTDKILPVETDKKPOGKQLQTRADYLLKLLRKGLEK-----GAVT 1332	
Qy	1148 GNNNASADGAQVNSMFYVRDMQRRLLVEFKGRVLLLEKAMNVEYAEYGLGSSSIPT 1207	
Db	1333 GGEEAK-----LKKRKPRVKENKVPRLKEEHGIELSSPRHSD 1370	
Qy	1208 EPEAEPKIADTVGSFIEVDDEMLDGLPKTDPTITSEEIMGAADVNNQARVEIAQHYNQMC 1267	
Db	1371 NPSEGEVKD-----DGLEKS-PMKKQKKKENKENKE-----KQMS 1406	
Qy	1268 KLLD-ENARESQAAYVNNQSPSTKVNESFRALKSINGNINTILSITSDQSKSHEDDTKPD 1326	
Db	1407 SRKDKEGDKERKSKDKKKEPKSGDAKSSSKRSQSPVHITAGSEVPVIGEDEDD---D 1463	
Qy	1327 LNNVEMKDTAEETKPLRGVVDLNVVEGEENIAE 1360	
Db	1464 LDQETFSICKERMRPVKKALKQLDKPKDGLNVQE 1497	
RESULT 7		
CHD1_MOUSE		
ID	CHD1_MOUSE	STANDARD; PRT; 1711 AA.
AC	P40201;	
DT	01-FEB-1995 (Rel. 31, Created)	
DT	01-FEB-1996 (Rel. 33, Last sequence update)	
DT	15-MAR-2004 (Rel. 43, Last annotation update)	
DE	Chromodomain-helicase-DNA-binding protein 1 (CHD-1).	
GN	CHD1 OR CHD-1.	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	

```
OX NCBI_TaxID=10090;
RN [1] SEQUENCE FROM N.A.
RP MEDLINE=932111972; PubMed=8460153;
RA Delmas V., Stokes D.G., Perry R.P.;
RT "A mammalian DNA-binding protein that contains a chromodomain and an
RT SNF2/SWI2-like helicase domain.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:2414-2418(1993).
CC -!- FUNCTION: Sequence-selective DNA-binding protein. Could play an
CC important role in gene regulation.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: Abundance is higher in cells representing
CC early stages of the B lymphoid lineage such as pre-B and B cells,
CC than in cells representing mature plasmacytes or other cell
CC lineages such as fibroblasts.
CC -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.
CC -!- SIMILARITY: Contains 2 chromo domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; L10410; AAB08486.1; -.
CC HSSP; P23197; IAP0.
CC MGD; MGI:88393; Chd1.
CC GO; GO:0005634; C:nucleus; NAS.
CC GO; GO:0003682; F:chromatin binding; NAS.
CC GO; GO:0004386; F:helicase activity; NAS.
CC GO; GO:0006333; P:chromatin assembly/disassembly; NAS.
CC InterPro; IPR000953; Chromo.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR001650; Helicase_C.
CC InterPro; IPR000330; SNF2_N.
CC Pfam; PF00385; chromo; 2.
CC Pfam; PF00271; helicase_C; 1.
CC Pfam; PF00176; SNF2_N; 1.
CC SMART; SM00298; CHROMO; 2.
CC SMART; SM00487; DEXDC; 1.
CC SMART; SM00490; HELICC; 1.
CC PROSITE; PS00598; CHROMO_1; 2.
CC PROSITE; PS00013; CHROMO_2; 2.
CC PROSITE; PS00690; DEAH_ATP_HELICASE; FALSE NEG.
KW Hydrolase; Helicase; Nuclear protein; ATP-binding; DNA-binding;
KW Repeat.
FT DOMAIN 1 70 SER-RICH.
FT DOMAIN 116 136 SER-RICH.
FT DOMAIN 270 362 CHROMO 1.
FT DOMAIN 387 450 CHROMO 2.
FT NP_BIND 504 511 ATP (POTENTIAL).
FT SITE 612 615 DEAH BOX.
FT DOMAIN 1629 1645 3 X 5 AA REPEATS OF H-S-D-H-R.
FT REPEAT 1629 1633 1.
FT REPEAT 1635 1639 2.
FT REPEAT 1641 1645 3.
SQ SEQUENCE 1711 AA; 196409 MW; FE3F8F8D13E32E24 CRC64;

Query Match 21.9%; Score 1572.5; DB 1; Length 1711;
Best Local Similarity 30.7%; Pred. No. 4.6e-72;
Matches 427; Conservative 239; Mismatches 410; Indels 313; Gaps 46;

QY 14 RKPVYNLDDSDDDFVPKKDRTFEQVEAIVRTDAKE-----NAQACGESTNLVSC 64
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 209 KRQIDSSSEDEDEDY--DNDKRSRRQATVNVSYKEDDEEMKTDSDLLLEVCGED----- 260

QY 65 NTCTYAFHAKCLVPPLKDAVENWRCPECVSPLNEIDKILDCEMRPTKSQGS----- 118
   || : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 261 -----VPQPEDEEFET-----IERVMDC--RVGRKATGATTTIYA 294

QY 119 --SDAEPKPIFVK-----QYLVKWKGLSYLHCSWVPEKEFQKAYKSNHRLKTRVNNFH 169
```

```
Db 295 VEADGDPNAGFERNKEPGDIQYLIKWKGWSHIHNTWTEETLK---QQNVRGMKKLDNYK 351
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 170 RQ-----MESFNSEDDFVAIRPEWTTVDRIACREED---GELEYLVKY 211
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 352 KKDQETKRWLKNASPEDVEYNCQOELTDDLHKQYQIVERIIAHSNQSAAGLPDYICKW 411
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 212 KELSDECYWESESDIS-TFONEIQRFKDVN-SRTRRSKDVD-HKRNPRDFQQFDHTPEF 268
   : | | | | | : | | : | : | : | : | : | : | : | : | : | : |
Db 412 QGLPYSECSWEDGALISKKFTCIDEIFSRNQSKTTPFKDCKVKLKQRP-RFVALKKQPSY 470
   : | | | | | : | | : | : | : | : | : | : | : | : | : | : |
QY 269 LKG----LLHPYQLEGLNFRFSWSKQTHVILADEMGLKTIQSIALLASLFEENLI--P 322
   : | | | | | : | | : | : | : | : | : | : | : | : | : | : |
Db 471 IGGHEGLELDYQLNGLNLWLAHSWCKGNSCILADEMGLGKTIQTISFLNYLFHEHQLYGP 530
   : : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
QY 323 HLVIAPLSTLRNWEREFATWAPQMNVMYFGTAQARAVIREHEFYLSKDQKIKKKKSGQ 382
   : : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 531 FLLVVPLSTLTQREIQTWASQMNNAVYLGDINSRNMIRTHEW----- 574
   : : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
QY 383 ISSSEKQKRIKFDVLLTSYEMINLDSAVLKPIKWECEMIVDEGHRLLKNKDSKLFSSLTQYS 442
   : : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 575 --MHPQTKRLKFNILLTYYEILLKDKAFLGGLNWAFIGVDEAHLKNDDSLKYKTLIDFK 632
   : : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
QY 443 SNHRIILTGTPLQNNLDELFMMLMHFLDAGKFGSLEEFQEEFKDINQEEQISRLHMKLAPH 502
   : : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 633 SNHRLITGTPLQNSLKELWSLLHFIMPEKFSWEDFEEHKG-GREYGYASLHKELEPF 691
   : : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
QY 503 LLRRVKDVMKDMPPKKELIRVDLSSSQKEYYKAIFTRNYQVLTK-KGGAQISLNNIMM 561
   : : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 692 LLRRVKDVEKSLPAKVEQILRMEMSALQYQYKWILTRNYKALSKGSGSTSGFLNIMM 751
   : : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
QY 562 ELRKVCCHPYMLEGVE-PVIHDANEAFKQLLESCGLQLLDKMMVKLKEQGHRLVIYTFQF 620
   : : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 752 ELKCCNHCYLIKPPDNNEFYNQEAQLHLSRSGKLLLDKLLIRLRERGNRVLIFSQM 811
   : : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
QY 621 QHMLDLEDYCTHKKQYERIDGKVGGAERQIRIDRFNAKNSNKFCLLSTRAGGLGINL 680
   : : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 812 VRMLDILAELYKYRQFPFQRLDGSIKGELRKQALDHFNAEGSEDFCFLSTRAGGLGINL 871
   : : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
QY 681 ATADTVIIYDSDNPHADLOAMARAHRLGQTNKVMYIYLRNGTIEERMQLTKKKMVL 740
   : : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 872 ASADTVVIFDSDNPNQNDLOAQAARAHRIQKKQVNIYRLVTGKSVVEEDILERAKKKMVLD 931
   : : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
QY 741 HLVGKLTQN-----INQELDDIIRYKSKELFASDEDEAGSKGKIHYD 785
   : : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 932 HLVIQRMDDTGKTVLHTGSAFSSSTPFNKEELSAILKFGABELFKEPEGEQEPEP-----Q 986
   : : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
QY 786 DAAIDKLLDRDLVEAEE---VSVDDDEENGFLKAFKANFEYIDENEAALAEQORVAES 842
   : : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 987 EMDIDELKRAETHENEPGLSVGDE---LLSQFKVANFSNMDEDD-IELEPER---NS 1038
   : : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
QY 843 KSSAGNSDRASYWEEL-----LKDKFELHQAELNALGKRKRKRKQLVSIIEEDDLA 894
   : : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1039 KN-----WEEIIPEEQRRLEEEERQKELEIYMLPRMNCAGQ----- 1077
   : : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
QY 895 LEDVSSDGDESYEAESTDGEAAGQVQTGRRPYRRKGRDNLPTPLMEGEGRSFRVLGFN 954
   : : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1078 ---ISFNGSEGRSRSRRYSGSDSDSISERKRPKKRGRPTIPRE-----NIKGF 1125
   : : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
QY 955 QSQRALFVQTLRMRYGAGNDFWKEFVPR---LKQKTFEEINEYGIILFKHIAEEIDENS-- 1009
   : : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1126 DAEIRRFIKSYKKFG-GPLERLDAIARDAELVDKSETDLRRLGELVHNGCVKALKDSSSG 1184
   : : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
QY 1010 -----PTFSDGVPKBGLRIEDVLRIALLILVQEKVKFVEDHPGKPVFPSPRI 1056
   : : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1185 TERAGRLGKVGKPTF-----RISGVQVNAKLVIHAHEDEL--IPLHKSIPSDPE-- 1231
   : : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
QY 1057 LERFPGRLRSGKI-----WKEEHDKIMIRAVLKHGYGRWQAIVDDKELGIQELI 1104
   : : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1232 -ER----KQYTIPTCHTKAAHFDIDWGKEDDSNLLIGIYGYGSWEMIKMDPDLSLTHKI 1286
   : : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
QY 1105 -----CKELNFPH---ISLSAAEQAGLQGN--GSGGSPGAQTNPNGSVITGN 1149
   : : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
```



Db 1287 LPDDPKKPOAKQLQTRADYLIKLSRDLAKREAQRLCGAGS----- 1329

QY 1150 NNASADGAQVNSMFYYRDMQRRLVEFVKRVLILLEKAMNVEYAEYVYGLGSSSIPTTEP 1209

Db 1330 -----KRRKTRAKSKAM-----KSIVKKEI--KSDSSPLPSEKS 1363

QY 1210 EAEPKIADTVGVSFIEVDDEMLDGLPKTDPITSEIM-----GAAVDNNQARVEI 1259

Db 1364 D-----EDDDKLNDSKPSKDRSKSVSDAPVHITASGEPVPIAESEEL 1409

QY 1260 AQHYNQMCK 1268

Db 1410 DQKTFISICK 1418

RESULT 8

CHD1 HUMAN

ID -CHD1\_HUMAN STANDARD; PRT; 1709 AA.

AC O14646;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Chromodomain-helicase-DNA-binding protein 1 (CHD-1).

GN CHD1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=97470991; PubMed=9326634;

RA Woodage T., Basrai M.A., Baxevarnis A.D., Hieter P., Collins F.S.;

RT "Characterization of the CHD family of proteins.";

RL Proc. Natl. Acad. Sci. U.S.A. 94:11472-11477(1997).

CC -!- FUNCTION: Sequence-selective DNA-binding protein. Could play an

CC important role in gene regulation.

CC -!- SUBCELLULAR LOCATION: Nuclear.

CC -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.

CC -!- SIMILARITY: Contains 2 chromo domains.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AF006513; AAB87381.1; -.

DR HSSP; P23197; IAP0.

DR Genew; HGNC:1915; CHD1.

DR MIM; 602118; -.

DR GO; GO:0004003; F:ATP dependent DNA helicase activity; TAS.

DR GO; GO:0007001; P:chromosome organization and biogenesis (sen. . .; TAS.

DR GO; GO:0006357; P:regulation of transcription from Pol II pro. . .; TAS.

DR InterPro; IPR000953; Chromo.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR001650; Helicase\_C.

DR InterPro; IPR000330; SNF2\_N.

DR Pfam; PF00385; chromo; 2.

DR Pfam; PF00271; helicase\_C; 1.

DR Pfam; PF00176; SNF2\_N; 1.

DR SMART; SM00298; CHROMO; 2.

DR SMART; SM00487; DEXDc; 1.

DR SMART; SM00490; HELICc; 1.

DR PROSITE; PS00598; CHROMO\_1; 2.

DR PROSITE; PS50013; CHROMO\_2; 2.

DR PROSITE; PS00690; DEAD\_ATP\_HELICASE; FALSE NEG.

KW Hydrolase; Helicase; Nuclear protein; ATP-binding; DNA-binding;

KW Repeat.

FT DOMAIN 1 70 SER-RICH.

FT DOMAIN 117 137 SER-RICH.

FT DOMAIN 272 364 CHROMO 1.

FT DOMAIN 389 452 CHROMO 2.

FT NP\_BIND 506 513 ATP (POTENTIAL).

FT SITE 614 617 DEAH BOX.

FT DOMAIN 1628 1644 3 X 5 AA REPEATS OF H-S-D-H-R.

FT REPEAT 1628 1632 1.

FT REPEAT 1634 1638 2.

FT REPEAT 1640 1644 3.

SQ SEQUENCE 1709 AA; 196517 MW; 416409C913D6A935 CRC64;

Query Match 21.7%; Score 1562.5; DB 1; Length 1709;

Best Local Similarity 32.4%; Pred. No. 1.5e-71;

Matches 410; Conservative 220; Mismatches 384; Indels 251; Gaps 41;

QY 9 RIRSDRKPVYNLDDDDDFVPKKDRTFEQVEAIVRTDAKE-----NACQACGEST 59

Db 204 KILGQKKRQIDSSEEDDEEDYDNDKRSSRRQATVNVSYKDEEMKTDSDDLLEVCGE- 262

QY 60 NLVSCNTCTYAFHAKCLVPPLKDAVENWRCPECVSNFLNEIDKILDCEM-----RPTKS 113

Db 263 -----VPQPEEEFET-----IERFMDCRIGRKGGATGATT 293

QY 114 SEQSSDAEPKPIFVK-----QYLVKWKGLSYLHCSWVPEKEFQKAYKSNHRLKTRVN 166

Db 294 IYAVEADGPNAGFEKNKEPGEIQYLIKWKGWSHIHTWETEETLK---QQNVRGMKKLD 350

QY 167 NFHRQ-----MESFNNSDDFVAIRPEWTTVDRI LACREED---GELEYL 208

Db 351 NYKKKQETKRWLKNASPEDVYYNCQQLTDDLHKQYQIVGRIIAHSNQKSAAGYPDY 410

QY 209 VKYKELSYDECYWESESDIS-TFQNEIQRFKDVN-SRTRRSKDVD-HKRNPRDFQQFDHT 265

Db 411 CKWQGLPYSECSWEDGALISKKFQACIDEYFSRNQSKTTPFKDCKVLKQRP-FVALKKQ 469

QY 266 PEFLKG---LLHPYQLEGLNFLRFSWSKQTHVILADEMGLKTIQSIALLASLFEENLI 321

Db 470 PSYIGGHEGLELDYQLNGLNWLASHWCKGNSCILADEMGLKTIQTSFLNYLFHHEQL 529

QY 322 --PHLVIAPLSTLRNWEREFATWAPQMNVMYFGTAQARAVIREHEFYLSKDQKKIKKK 379

Db 530 YGPFLLVVPLSTLTSWOREIQTWASQMNNAVYLGDINSRNMI RTHEW----- 576

QY 380 SGQISSESKOKRIKFDVLLTSYEMINLDSAVLKPIKWECMIVDEGHRLLKNKDSLFSSLT 439

Db 577 -----THQTKRLKFNILLTYEILLKDKAFLGGLNWAFIGVDEAHLKNDDSLKYTLI 631

QY 440 QYSSNHRILLTGTPLONNIDELFMLMHFLDAGKFGSLBEFQEEFKDINQEEQISRLHKML 499

Db 632 DFKSNHRLLLITGTPLONSLKEMLWSLLHFIMPEKFSWEDFEEHKG-GREYGYASLHKEL 690

QY 500 APHLLRRVKDKVMKMPKKELILRVDLSLQKEYYKAIFTRNYQVLT-K-KGGAQISLNN 558

Db 691 EPFLLRVKDKVEKSLPAKVEQILRMEMSALQKQYKWLTRNYKALSKGSGSTSGFLN 750

QY 559 IMMELRKVCCHPYMLEGVE-PVIHDANEAFKQLLESCGKLQLLDKMMVKLKEQGHVLIY 617

Db 751 IMMELKKCCNHCYLIKPPDNNEFYNKQEAQLHLIRSSGKLILLDKLLIRLRERGNRVLI 810

QY 618 TQFQHMLDLEDYCTHKKWQYERIDGKVGGAEQRIIDRFNAKNSNKFCLLSTRAGLG 677

Db 811 SQMVRMLDILAELYKYRQFPFQRLDGSIKGELRKQALDHFNAEGSEDFCLLSTRAGLG 870

QY 678 INLATADTVIYDSDWNPHADLQAMARAHRLGQTNKVMYRLINRGTI EERMQLTKKM 737

Db 871 INLASADTVVIFDSWNPQNDLQAQARAHRIQKQKQVNIYRLVTKGSVEEDILERA KKM 930

QY 738 VLEHLVVGKLTQN-----INQEELDDIIRYGSKELFASEDD EAGSGKI 782

Db 931 VLDHLVIQRMDDTTGKTVLHTGSAPSSSTPFNKEELSAILKFGAEELFKEPEGE EOEPE--- 987

QY 783 HYDDAAIDKLLDRDLVEAEE---VSDDEEENGFLKAFKVFANFEYIDENEAAL EQRVA 839

Db 988 --QEMDIDEILKRAETHENEPGPLTVGDE----LLSQPKVANFSNMD EDD-IELEPER-- 1038

QY 840 AESSAGNSDRASYWEEL-----LKDKFELHQAEEALNALGKKRKRKQLVSIEEDD 891  
Db 1039 -NSKN-----WEEIIPEDQRRRLLEEERQKELEEIYMLPRMNCACQ----- 1079  
QY 892 LAGLEDVSSGDDES YEAE-----TDGEAAGQGVQTGRRPYRRKGRDNLEPTPLMEGEGR 946  
Db 1080 -----ISFNGSEGRRSRRYSGSDSDS-----ISEGKRP-KKGRPRTIPRE----- 1121  
QY 947 SFRVLGFNQSORAIFVQTLRMRYGAGNEDWKEFVPR---LKQKTFEEINEYIGILFLKHIAE 1003  
Db 1122 --NIKGFSDAEIRRFIKSYKKFG-GPLERLDAIARDAELVDKSEFIDLRLRGELVHNGCIK 1178  
QY 1004 EIDENS-----PTFSDGVPKEGLRIEDVLVRIALLILVQEKVKFVEDHFG 1048  
Db 1179 ALKOSSSGTERTGGRLGKVKGPTF-----RISGVQVNAKLVISHEEEL--IPLHKS 1227  
QY 1049 KPVFPPSRILERPPLRSGKI-----WKEHDKIMIRAVLKHGYGRWQAIVDDK 1096  
Db 1228 IPSDPE---ER-----KQYTIPTCHTKAAHFIDWKGEDDSNLLIGIYEGYGSWEMIKMDP 1280  
QY 1097 ELGIQELI-----CKELN-----FPHISLSAAEQAGLQGGSGGSPGQATNQ 1140  
Db 1281 DLSLTHKILPDDPKKPPQAKQLQTRADYLIKLLSRDLAKKEALSGAGSSKRRKARAKNK 1340  
QY 1141 NPGSV 1145  
Db 1341 AMKSI 1345

RESULT 9  
CHDB\_HUMAN  
ID CHDB\_HUMAN STANDARD; PRT: 2004 AA.  
AC Q9HCK8; Q8N3Z9; Q8NCY4; Q8TBR9; Q96F26;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Chromodomain-helicase-DNA-binding protein HELSNF1 (Helicase with SNF2 domain 1) (Fragment).  
GN HELSNF1 OR KIAA1564.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=22158633; PubMed=12168954;  
RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;  
RT "Construction of expression-ready cDNA clones for KIAA genes: manual curation of 330 KIAA cDNA clones."  
RL DNA Res. 9:99-106(2002).  
RN [2]  
RP SEQUENCE OF 588-2004 FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=20450683; PubMed=10997877;  
RA Nagase T., Kikuno R., Nakayama M., Hirose M., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XVII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro."  
RL DNA Res. 7:273-281(2000).  
RN [3]  
RP SEQUENCE OF 1554-2004 FROM N.A.  
RC TISSUE=Brain, and Lung;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [4]  
RP SEQUENCE OF 1768-2004 FROM N.A.  
RC TISSUE=Lymph node;  
RA Dueterhoeft A., Lauber J., Mewes H.-W., Weil B., Wiemann S.;  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Probable transcription regulator.  
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.  
CC -!- SIMILARITY: Contains 2 chromo domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; AB046784; BAB13390.2; -.  
DR EMBL; BC011695; AAH11695.1; ALT\_INIT.  
DR EMBL; BC025964; AAH25964.1; -.  
DR EMBL; BC036920; AAH36920.1; -.  
DR EMBL; AL834524; CAD39180.1; -.  
DR Genew; HGNC:20153; HELSNF1.  
DR GO; GO:0005634; C:nucleus; NAS.  
DR GO; GO:0030528; F:transcription regulator activity; NAS.  
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.  
DR InterPro; IPR006576; BRK.  
DR InterPro; IPR000953; Chromo.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR001650; Helicase\_C.  
DR InterPro; IPR000330; SNF2\_N.  
DR Pfam; PF00271; helicase\_C; 1.  
DR Pfam; PF00176; SNF2\_N; 1.  
DR SMART; SM00592; BRK; 2.  
DR SMART; SM00298; CHROMO; 2.  
DR SMART; SM00487; DEXDc; 1.  
DR SMART; SM00490; HELICc; 1.  
DR PROSITE; PS00598; CHROMO\_1; FALSE\_NEG.  
DR PROSITE; PS50013; CHROMO\_2; 1.  
DR PROSITE; PS00690; DEAH\_ATP\_HELICASE; FALSE\_NEG.  
KW Transcription regulation; Hydrolase; Helicase; Chromatin regulator;  
KW Nuclear protein; ATP-binding; DNA-binding; Repeat.  
FT NON\_TER 1  
FT DOMAIN 65 132 CHROMO 1.  
FT DOMAIN 147 213 CHROMO 2.  
FT NP\_BIND 259 266 ATP (POTENTIAL).  
FT SITE 371 374 DEAH\_BOX.  
FT DOMAIN 1492 1521 SER-RICH.  
FT DOMAIN 1916 1931 HIS-RICH.  
FT DOMAIN 1962 2004 ASP-RICH.  
FT CONFLICT 1991 1991 M -> I (IN REF. 3; AAH36920).  
SQ SEQUENCE 2004 AA; 230356 MW; 811124364AF791DF CRC64;  
  
Query Match 20.8%; Score 1494; DB 1; Length 2004;  
Best Local Similarity 31.7%; Pred. No. 5.5e-68;  
Matches 408; Conservative 214; Mismatches 419; Indels 246; Gaps 45;  
  
QY 78 PPLKDASVENWRCPECVSPLEIDKILDCEMRPTKSSEQSSDAEPKPIFVKQYLVKWK 137  
Db 59 PSEEDAAI-----VDKVL--SMRIVKELPSGQYTE-----AEEFFVKYN 97

QY 138 LSYLHCSWVPEKEFKAYKSNHRLKTRVNNFHRQMESFNNSDDFAIRPEWTTVDRII- 196  
Db 98 YSYLHCEWATISQLEKDKRIHQKLRFKTKMAQMRHFFHEDEEPF---NPDYVEVDRIID 154  
QY 197 ---ACREEDGE--LEYLVKYKELSYDECYWESESDISTFQNEIQPFKDVNSRTRRSKQVD 251  
Db 155 ESHSIDKNGEPVIYLVKWCSLPYEDSTWELKEDVD--EGKIRFPKRIQSRHPCLKRV- 211  
QY 252 HKRNPDRFQPDHTPEFL-KGLLHPYQLEGLNFLRFSWSKQTHVILADEMGLKTIQISIA 310  
Db 212 NRPQASAWKKLELSHEYKRNQRLREYQLEGVNWLFWNWRNQCILADEMGLKTIQISIA 271  
QY 311 LLASLFEENLI-PHLVIAPLSTLRNWEREFATWAPQMNVMVYFGTAQARAVIREHEFYLS 369  
Db 272 FLQEVNVGIGHPFPLVIAPLSTITNWEREBFTWT-EMNTIVYHGSLASRQMIQQYEMYC- 329  
QY 370 KDQKKIKKKSQGISSESQKRIKFDVLLTSYEMINLDSAVLKPIKWECMIVDEGHRKLN 429  
Db 330 -----KDSGRGLIPGA---YKFDALITTFEMILSDCPLEIEWRWCVIIIDEAHLKLN 378  
QY 430 KDSKLFSSLTQYSSNHRILLTGTPLONNLDELMLHFLDAGKFGSLEEFQEEFKDINQE 489  
Db 379 RNCKLLDSLKHMDLEHKVLLTGTPLONTVEELFSLHFLPSQFPSESEFLKDFGLKTE 438  
QY 490 EQISRLHKMLAPHLRLRVKQDVMDMPKKELILRVDLSSLOKEYYKAIFTRNYQVLTCK 549  
Db 439 EQVQKLAQLKPMMLRLKEDVEKNLAPKQETIIIEVELTNIQKYRAILEKFNFSFLSKG 498  
QY 550 GG--AQISLNNIMMELRKVCCHPYMLEGVEPVI-----HDANEAFKQLLESCG 595  
Db 499 AGHTNMPNLLNTMELRKCCNHPYLINGAEKILTEFREACHIIPHDFH--LQAMVRSAG 556  
QY 596 KLQLDKMMVKLKEQHRVLIYTOFQHMDDLLEDYCTHKWKWVERIDGKVGGAERQIRID 655  
Db 557 KLVLIDKLLPKLKAGGHKVLIFSQMVRCILDILEDYLIQRRYLYERIDGRVGNLRQAAD 616  
QY 656 RFNAKNSKFCPLLSTRAGGLGINLATADTVIYDSDNPHADLOAMARHRLGQTNKVM 715  
Db 617 RFSKPDSDRFVFLCTRAGGLGINLTAADTCIIFSDWNPNQNDLQAQARCHRIGQSKAVK 676  
QY 716 IYRLINRGTTIERMMLQTKKMWLEHLVVGKLGK-----TQINQBEELDDIIRYGSKE 767  
Db 677 VYRLITRNSVEREMFDFKASLKLGLDKAVLQSMGRDGNITGIIQFSPKKEIEDLLRKGAYA 736  
QY 768 LFASEDEACKSGKIHYYDDAAIDKLLDRDLVEAEVSVDDDEENGFLKAFKVANFEYIDE 827  
Db 737 AIMEEDDEGSK-----FCEEDIDQIL---LRRTTITIESE---CKGSTFAKASF----- 780  
QY 828 NEAAALEAQRVAESKSSAGNSDRASYWEELLKDKFELHQAEELNALGKRK-----RS 880  
Db 781 -----VASENRTDI-SLDDPNFWQKWAK-KADL----DMDLLNSKNNLVIDTPRV 824  
QY 881 RKQ---LVSIEDDDLAGLEDVSSDGDSEYAEASTDGEAAGQGVQTGRPPYRRKGRDNLEP 937  
Db 825 RKQTRHFSLLKDDDLVEFSDLESEDDDERPRRRHRHA-----YGRTDCCFRVEK 874  
QY 938 TPLMEGEGRSFRVLGFNQSQ-----RAIFVQTLTRY-GAGN---FDWKEFVPR 981  
Db 875 HLLVYGWGRWRDILSHGRFKRRMTERDVETICRAILVYCLLHVRGDNENIKGFIDWLISPA 934  
QY 982 LKQKTFFEEINEYGI-----LFLKHIAEEIDENSP--TFSDGVPKEGLR 1022  
Db 935 ENGKTKELQNSGLSIPVPRGRKGKVKKSQSTFDIHKADWIRKYNPDTLFQDESYKXHLK 994  
QY 1023 --IEDVLVRIALLILVQEKVKFVEDHPGKPV---FPSRILERPGLRSGKI---WKEEH 1073  
Db 995 HQCNKVLRLVRMLYLRQEV--IGDQAEKVLGGAIASEIDIWFPVVDQLEVPVTTWWDSEA 1052  
QY 1074 DKIMIRAVLKHGVRWQAI-----VDDKELGIGELICKELNFPPI----- 1113  
Db 1053 DKSLLIGVFKHGYEKYNTMRADPALCFLEKAGRPDDKAIAAEHRVLD--NFSDIVEGVDF 1110  
QY 1114 --SLSAAEQAQGLQG-----QNGSGG-----SNPGAQTNQNPGSVITGNNASAD 1155

Db 1111 DKCEDPEYKPLQPPKQDDDEGDLMMDEEISVIDGDEAQTQOPGHLFWPPGSALTA 1170  
QY 1156 GAQVNSMFYYR-----DMQRRIVE--FVKRVLLEKAMNVEYAE--EY 1196  
Db 1171 RLRLVTAYQRSYKREQMKIEAERDRRRRCEAAPFKLKEIARREKQQRWTRREQTDFY 1230  
QY 1197 GLGSSSSIPTEEPEAEPKIADTVGVSF 1223  
Db 1231 -----RVVSTFGVEY 1240  
RESULT 10  
CHDA HUMAN  
ID CHDA HUMAN STANDARD; PRT; 1967 AA.  
AC Q9P2D1;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Probable chromodomain-helicase-DNA-binding protein KIAA1416  
DE (Fragment).  
GN KIAA1416.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=20181126; PubMed=10718198;  
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XVI.  
RT The complete sequences of 150 new cDNA clones from brain which code  
RT for large proteins in vitro.";  
RL DNA Res. 7:65-73(2000).  
CC -!- FUNCTION: Probable transcription regulator.  
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.  
CC -!- SIMILARITY: Contains 2 chromo domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AB037837; BAA92654.1; -.  
DR InterPro; IPR006576; BRK.  
DR InterPro; IPR000953; Chromo.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR001650; Helicase C.  
DR InterPro; IPR001005; Myb\_DNA\_Binding.  
DR InterPro; IPR000330; SNF2\_N.  
DR Pfam; PF00271; helicase\_C; 1.  
DR Pfam; PF00176; SNF2\_N; 1.  
DR SMART; SM00592; BRK; 2.  
DR SMART; SM00298; CHROMO; 2.  
DR SMART; SM00487; DEXDC; 1.  
DR SMART; SM00490; HELICc; 1.  
DR SMART; SM00717; SANT; 1.  
DR PROSITE; PS00598; CHROMO\_1; FALSE\_NEG.  
DR PROSITE; PS50013; CHROMO\_2; 2.  
DR PROSITE; PS00690; DEAH\_ATP\_HELICASE; FALSE\_NEG.  
KW Transcription regulation; Hydrolyase; Helicase; Chromatin regulator;  
KW Nuclear protein; ATP-binding; DNA-binding; Repeat.  
FT NON\_TER 1  
FT DOMAIN 38 105 CHROMO 1.  
FT DOMAIN 120 185 CHROMO 2.  
FT NP\_BIND 231 238 ATP (POTENTIAL).  
FT SITE 343 346 DEAH BOX.  
FT DOMAIN 1177 1183 POLY-ARG.



FT	DOMAIN	1403	1496	GLU-RICH.
FT	DOMAIN	1636	1643	POLY-ARG.
FT	NON TER	1967	1967	
SQ	SEQUENCE	1967 AA;	225722 MW;	4C13FC4E4A1EBB15 CRC64;
Query Match				
Best Local Similarity		20.8%;	Score 1492.5;	DB 1; Length 1967;
Matches 407;		Conservative 228;	Mismatches 423;	Indels 279; Gaps 48;
Qy	100	IDKILDCEMRPTKSSEQSSDAEPKIPVQYLVKWKGLSYLHCSWVPEKEFQKAYKSNH	159	
Db	40	VEKIM--SSRSVKQKESGEEVE-----IEEFYVKYKNFSYLHCQWASIEDLEK----	DK 88	
Qy	160	RLKTRVNNF---HRQMESFNNSDEDFVAIRPEWTTVDRIIL---ACREEDGE--LEYLVKY	211	
Db	89	RIQQIKIRFKAQGNKFLSEIEDEL--FNPDYVEVDRIIMDFARSTDDRGEVPVTHYLVKW	146	
Qy	212	KELSYDECYWESESISTFQNEIQRFKDVNSRTRRSKVDHKKRNP-RFQQFDHTPEFL-	269	
Db	147	CSLPYEDSTWERRQDID--QAKIEEFKELMSREPETERVE--RPPADDWKKSSESSREYKN	202	
Qy	270	KGLLHPYQLEGLNFLRFSWSKQTHVILADEMGLKTIQSIALLASLFEENLI-PHLVIAP	328	
Db	203	NNKLREYQLEGVNWLLFNWYNMRNCILADEMGLKTIQSIITFLYIYLKGIHGPFVLIAP	262	
Qy	329	LSTLRNWEREATWAPQMNVMYFGTAQARAVIREHEFYLSKDQKIKKKKSGQISSSEK	388	
Db	263	LSTIPNWEREFRTWT-ELNVVVYHGSQASRRTIQLYEMYFKDPQGRVIKGS-----	312	
Qy	389	QKRIKFDVLTSYEMINLDSAVLKPWKWECMIVDEGHRLLKNKDSKLFSSLTQYSSNHRIL	448	
Db	313	---YKFHAIIITTFEMILTDCEPRLNIPWRCWVIDEAHRLKNRNCKLLEGLKXMDLEHKVL	369	
Qy	449	LTGTPLQNLLDELFMLMHFDAGKFGSLBEFQEEFKDINQEEQISRLHKMLAPHLRRVK	508	
Db	370	LTGTPLQNTVEELFSLHLFLEPSRFPSETFMQEFGLKTEEQVKLQAILKPMMLRLK	429	
Qy	509	KDVMKDMPPKKELILRVDLSSLQKEYKAIFTRNYQVLTKKGG-AQI-SLNNIMMELRKV	566	
Db	430	EDVEKNLAPKEETIEVELTNIQKKYRAILEKNFTFLSKGGGQANVPNLLNTMMELRKC	489	
Qy	567	CCHPYMLEGVPEVI-----HDANEAFKQL---LESCGKLQLLDKMMVKLKEQGHRVLI	616	
Db	490	CNHPYLINGAEEKILEEFKETHNAESPDLQAMIQAAGKVLIDKLLPKLKAGGHRVLI	549	
Qy	617	YTQFQHMLDLLEDYCTHKKWQYERIDGKVGAERQIRIDRFNAKNSNKFCELLSTRAGGL	676	
Db	550	FSQMVRCLDILEDYLIQRRYPYERIDGRVGNLRQAAIDRFSPKPSDRFVFLCTRAGGL	609	
Qy	677	GINLATADTVIYSDNPNHADLQAMARHRLGQTNKVMYRLINRGTIIEERMMLTKKK	736	
Db	610	GINLTAADTCIIFDSDNWPQNDLQAQARCHRIGQSKSVKIYRLITRNSYEREMFDKASLK	669	
Qy	737	MVLEHLVVGKLK-----TONINQEEELDDIIRYGSKELFASEDDEAGSKGIHYDDAA	788	
Db	670	LGLDKAVLQSMGRENATNGVQQLSKKEIETHLLRKGAYGALMDEDEGSK-----FCEED	724	
Qy	789	IDKLLDRDLVEAEVSVDDDEENGFLKAPKVANPEYIDENEAALAEQRVAESKSSAGN	848	
Db	725	IDQIL---LRRTHITIESE---GKGSTFAKASF-----VASGNRTDI-S	762	
Qy	849	SDRASWEELDKDFELHQAEELNALGKPK-----RSRKQ---LVSIEEDDLAGLEDV	898	
Db	763	LDDPNFWQWAK-KAEL---DIDALNGRNNLVIDTPRVRKQTRLYSAVKEDELMEFSDL	817	
Qy	899	SSDGDESYEASTDGEAAGQGVQTRRRPYRRKGRDNLEPTPLMEGEGRSFRLGNFSQ-	957	
Db	818	ESDSEKPCAKPRRPQDKSQG-----YARSECFRVEKNLLVYGWRWTDILSHGRYKR	870	
Qy	958	-----RAIFVQTLRMRY---GAGNFDWKEFVPRLKQKTFBEINEYGI-----	995	
Db	871	QLTEQDVETICRTILVYCLNHYKGDENIKSFIWDLITPTADGQTRALVNHSGLSAPVPRG	930	

QY	996	-----LFLKHIAEEID-----ENSPTFSDGVPKEGLR--IEDVLVRIALLILVQEKV	1040
Db	931	RKGKKVKAQSTQPVVQDADWLASCNPDALFQEDSYKXHLKHCHCNKVLRLVRMLYYLRQEV	990
QY	1041	-----KFVE--DHPGKPVFPSPRILRFPGLRSGKIWKKEHDKIMIRAVLKHGYGRWQAI	1092
Db	991	IGDQADKILEGADSSEADVW--IPEPFHAEVPADWWDKADKSLIGVFKHGYEKYNSM	1047
QY	1093	VDDKELGIQELICKELNFPHISLSAAEQAG--LQONGSGG-----	1131
Db	1048	RADPALCFLE---RVGMPDAKAIAAEQRGTMDLADGGDGEFDEDEPEYKPTRTPFK	1103
QY	1132	-----SNPGAQTNQNPGSVITGNNNASADGAQVNSMFY-----	1164
Db	1104	DEIDEFANSPSEDKESMETHATGKHSES--NAELGQLYWPNTSTLTTLRRLRLITAYQRS	1161
QY	1165	YRDMQRRLVEFVK-----KRVLLEKAMNYEYAE--YYGLGGSSSI	1204
Db	1162	YKQQMRQEALMKTDRRRRRPREEVRALEAREAIISEKRQKWTREEADFY-----	1213
QY	1205	PTEEPEAEPKIADTVGSFIEVDDEMLDGLPKTDPITSEEIMGAADVNNQARVEIAQHYN	1264
Db	1214	-----RVVSTFGVIF-----DPVKQO-----FDWNQFRA-----	1237
QY	1265	QMCKLLDENARESVOAY	1281
Db	1238	--FARLDKKSDESLEKY	1252
RESULT 11			
HRP3_SCHPO			
ID	HRP3_SCHPO	STANDARD;	PRT; 1388 AA.
AC	O14139;		
DT	16-OCT-2001	(Rel. 40, Created)	
DT	16-OCT-2001	(Rel. 40, Last sequence update)	
DT	15-MAR-2004	(Rel. 43, Last annotation update)	
DE	Chromodomain helicase hrp3.		
GN	HRP3 OR SPAC3G6.01.		
OS	Schizosaccharomyces pombe (Fission yeast).		
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;		
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;		
OC	Schizosaccharomycetes.		
OX	NCBI_TaxID=4896;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=972;		
RX	MEDLINE=21848401; PubMed=11859360;		
RA	Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,		
RA	Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,		
RA	Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,		
RA	Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,		
RA	Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,		
RA	Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,		
RA	James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,		
RA	Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,		
RA	Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,		
RA	Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,		
RA	Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,		
RA	Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,		
RA	Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,		
RA	Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,		
RA	Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,		
RA	Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,		
RA	Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,		
RA	Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,		
RA	Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,		
RA	Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,		
RA	Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,		
RA	Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,		
RA	Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,		
RA	Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;		
RT	"The genome sequence of Schizosaccharomyces pombe.";		
RL	Nature 415:871-880(2002).		



Db 1228 RQSGKPNVKSQAIESATRTPTSPAISESRKKPSSKDTKIETPSREQRSQTASPVKSEK 1287  
QY 1321 DDTKPDNNVEMKDTAEETKPLR 1343  
Db 1288 DGNVSLNAEQKARCKELMYPVR 1310

RESULT 12  
CHD1 YEAST STANDARD; PRT; 1468 AA.  
AC P32657;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE CHD1 protein.  
GN CHD1 OR YER164W OR SYCP-ORF4.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288c / AB972;  
RX MEDLINE=97313264; PubMed=9169868;  
RA Dietrich F.S., Mulligan J.T., Hennessy K.M., Yelton M.A., Allen E.,  
RA Araujo R., Aviles E., Berno A., Brennan T., Carpenter J., Chen E.,  
RA Cherry J.M., Chung E., Duncan M., Guzman E., Hartzell G.,  
RA Hunicke-Smith S., Hyman R.W., Kayser A., Komp C., Lashkari D., Lew H.,  
RA Lin D., Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P.,  
RA Oh C., Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T.,  
RA Smith V., Taylor P., Wei Y., Botstein D., Davis R.W.;  
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome V.";  
RL Nature 387:78-81(1997).  
CC -|- SUBCELLULAR LOCATION: Nuclear (Potential).  
CC -|- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.  
CC -|- SIMILARITY: Contains 2 chromo domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U18917; AAB64691.1; -.  
DR PIR; S30818; S30818.  
DR GermOnline; 139241; -.  
DR SGD; S0000966; CHD1.  
DR GO; GO:0008023; C:transcription elongation factor complex; IPI.  
DR GO; GO:0016887; F:ATPase activity; IDA.  
DR GO; GO:0016944; F:Pol II transcription elongation factor acti. . ; IPI.  
DR GO; GO:0006338; P:chromatin modeling; IDA.  
DR GO; GO:0006368; P:RNA elongation from Pol II promoter; IPI.  
DR InterPro; IPR000953; Chromo.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR001650; Helicase\_C.  
DR InterPro; IPR000330; SNF2\_N.  
DR Pfam; PF00385; chromo; 2.  
DR Pfam; PF00271; helicase C; 1.  
DR Pfam; PF00249; myb\_DNA-Binding; 1.  
DR Pfam; PF00176; SNF2\_N; 1.  
DR SMART; SM00298; CHROMO; 2.  
DR SMART; SM00487; DEXDC; 1.  
DR SMART; SM00490; HELICC; 1.  
DR PROSITE; PS00598; CHROMO\_1; 2.  
DR PROSITE; PS50013; CHROMO\_2; 2.  
DR PROSITE; PS00690; DEAH\_ATP\_HELICASE; FALSE NEG.  
KW Hydrolase; Helicase; Nuclear protein; ATP-binding; DNA-binding;  
Repeat.  
FT DOMAIN 195 257 CHROMO 1.  
FT DOMAIN 285 350 CHROMO 2.

FT NP BIND 401 408 ATP (POTENTIAL).  
FT SITE 513 516 DEAH BOX.  
SQ SEQUENCE 1468 AA; 168240 MW; 78DB74C7FEC6BE5 CRC64;  
Query Match 20.3%; Score 1461.5; DB 1; Length 1468;  
Best Local Similarity 32.1%; Pred. No. 1.6e-66;  
Matches 410; Conservative 217; Mismatches 403; Indels 249; Gaps 46;  
QY 7 RLRIRSDRKPVYNLDDSDDDFVPKKDRTFEQVEAIVRTDAKENACQACGESTNLVSCNT 66  
Db 126 RFSNRQNKTVNYNIDYSDDDLLESEDD--YGSEAL-----SEENVHEA-----SANP 171  
QY 67 CTYAFHAKCLV--PPLKDASVENWRCPECVSPLEIDKILDCEMRPTKSSEQSSDAEPK 124  
Db 172 QPEDFHGIDIVINHRLKTSLEEGKVLEKTVPDLN-----NCKEN----- 210  
QY 125 PIFVKQYLVKWGLSYLHCSWVPEKEFQKAYKSNHRLK--TRVNNFHRQ-----ME 173  
Db 211 ----YEFLIKWTDESHLNTW-----ETYESIGQVRGLKRLDNYCKQFIIEDQQVRLD 259  
QY 174 SFNNSD-----DFVAIR-----PEWTTVDRIACRE----EDG--ELEYLVKYKELSYDEC 219  
Db 260 PYVTAEDIEIMDMERERRLDEFEFHVPERIIDSQRASLEDGTSQLQYLVKWRRLNYDEA 319  
QY 220 YWESSDISTFONE-IQRFKD-VNSRTRRSKQDVHKNRPDRDQQFDHTPEFLK-GLLHPY 276  
Db 320 TWENATDIVKLAPEQVKHFQRENISKILPOYSNNYTSQRPFEKLSVOPPFIKGGELRDF 379  
QY 277 QLEGLNFLRFSWSKQTHVILADEMGLGKTIQSIALLASLF--EENLPHLVIAPLSTLRN 334  
Db 380 QLTGINWMAFLWSKGDNGILADEMGLGKTVQTVAFISWLIPARRQNGPHIIVVPLSTMPA 439  
QY 335 WEREFAWAPQMNVMVYFGTAGARAVIREHEFYLSKDQKKIKKKSGQISSSESKQKRIKF 394  
Db 440 WLDTFEKWAPDLNCICYMGNOXSRDITIREYEFYTNPRAK-----GKTKMKF 485  
QY 395 DVLLTSYEMINLDSAVLKPIKWECMIVDEGHRLKNKDSKLFPSSLTOYSSNHRILLTGTP 454  
Db 486 NVLLTYEYILKDRAELGSIKWQFMAVDEAHLKNAESSLYESLSNFKVANRMLITGTP 545  
QY 455 QNNLDELFMHFLDAGKFGSLEEFQEEFKDINQEEQISRLHKMLAPHLLRRVKKDVMDK 514  
Db 546 QNNIKELAAALVNFMPGRFTIDQEIFENQDEEQEYIHDLHRRIQPFILRRLKKDVEKS 605  
QY 515 MPPKKEILIRVDLSSLQKEYYKAIETRNYQVLT--KKGGAQISLNNIMMELRKVCCHPYM 572  
Db 606 LPSKTERILRVELSDVQTEYYKNILTKNYSALTAGAKG-HFSLNIMNELKKNASHPYL 664  
QY 573 LEGVEPVI-----HDANAEAFKQLLESCGKLQLLDKMMVKLKEQGHRLVIYTFQHML 624  
Db 665 FDNAEERVLQKFGDGKMTRENVLRGLIMSSGKMVLDDQLTLRLKKGHRVLIFSQVMRL 724  
QY 625 DLLEDYCHKKWQYERIDGKVGGAERQIRIDRFNAKNSNKFCLLSTRAGGLGINLATAD 684  
Db 725 DILGDYLSIKGINFORLDGTVPQAQRISIDHFNSPDSNDFVLLSTRAGGLGINLMTAD 784  
QY 685 TVIYDSDWNPHADLOAMARAHRLGQTNKVMYIRLINRGTIIEERMVQLTKKKMVLHLVV 744  
Db 785 TVVIFDSDMNPQADLOAMARAHRIQGNHVMVYRLVSKDITVEEVLERARKMILEYAIL 844  
QY 745 G-----KLKTONINQEEELDDIIRYGSKELFASEDEAGSKGIHYDDA---AIDKLL 793  
Db 845 SLGVTGDNKYTKKNEPNAGELSAILKFGAGNMFTATDNQK-KLEDNLNDDVLNHAEDHVT 903  
QY 794 DRDLVEAEVSVDDDEENGFLKAFKVANF-----EYIDENEAALAEARVAAESKSSA 846  
Db 904 TPDLGESH--LGEE--FLKQFEVTDYKADIDWDDIIPPEELKKLQ----- 945  
QY 847 GNSDRASYWEELLKDKFELHQAEEELNALGKRKRSRKQLV-SIEDDDLGLAGLEDVSSDGDES 905  
Db 946 -----DEEQKRKDEEYVKEQLEMMNRRDNALKKIKNSVNGDGTA----ANSDD- 991  
QY 906 YEAESTDGEAAGQGVQTGRRPYRRKGRDNLPTPLMEGEGRSF--RVLGFNQ----- 955







RT "Isolation and characterization of hrpl+, a new member of the  
RT SNF2/SWI2 gene family from the fission yeast Schizosaccharomyces  
RT pombe.";  
RL Mol. Gen. Genet. 257:319-329(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RX MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
RT "The genome sequence of Schizosaccharomyces pombe.";  
RL Nature 415:871-880(2002).  
RN [3]  
RP SEQUENCE OF 852-1021 FROM N.A.  
RC STRAIN=968 h90;  
RX MEDLINE=20223868; PubMed=10759889;  
RA Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,  
RA Hiraoka Y.;  
RT "Large-scale screening of intracellular protein localization in living  
RT fission yeast cells by the use of a GFP-fusion genomic DNA library.";  
RL Genes Cells 5:169-190(2000).  
RN [4]  
RP CHARACTERIZATION.  
RX MEDLINE=20221599; PubMed=10756203;  
RA Yoo E.J., Jin Y.H., Jang Y.K., Bjerling P., Tabish M., Hong S.H.,  
RA Ekwall K., Park S.D.;  
RT "Fission yeast hrpl, a chromodomain ATPase, is required for proper  
RT chromosome segregation and its overexpression interferes with  
RT chromatin condensation.";  
RL Nucleic Acids Res. 28:2004-2011(2000).  
CC -!- FUNCTION: Seems to play a role in mitotic chromosome segregation  
CC and maintenance of chromatin structure. Has ATPase activity.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.  
CC -!- SIMILARITY: Contains 2 chromo domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X99021; CAA67494.1; -.  
DR EMBL; AL136499; CAB66168.1; -.  
DR EMBL; AB027852; BAA87156.1; -.  
DR PIR; T43334; T43334.  
DR PIR; T50107; T50107.  
DR GeneDB SPombe; SPAC1783.05; -.  
DR GO; GO:0005717; C:chromatin; ISS.

DR GO; GO:0005694; C:chromosome; ISS.  
DR GO; GO:0005654; C:nucleoplasm; ISS.  
DR GO; GO:0007049; P:cell cycle; ISS.  
DR GO; GO:0016043; P:cell organization and biogenesis; ISS.  
DR GO; GO:0006325; P:establishment and/or maintenance of chromatin.; ISS.  
DR GO; GO:0006350; P:transcription; ISS.  
DR InterPro; IPR000953; Chromo.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR001650; Helicase\_C.  
DR InterPro; IPR000330; SNF2\_N.  
DR Pfam; PF00385; chromo; 2.  
DR Pfam; PF00271; helicase\_C; 1.  
DR Pfam; PF00176; SNF2\_N; 1.  
DR SMART; SM00298; CHROMO; 2.  
DR SMART; SM00487; DEXDC; 1.  
DR SMART; SM00490; HELICC; 1.  
DR PROSITE; PS00598; CHROMO\_1; FALSE\_NEG.  
DR PROSITE; PS0013; CHROMO\_2; 2.  
DR PROSITE; PS00690; DEAH\_ATP\_HELICASE; FALSE\_NEG.  
KW Hydrolase; Helicase; Nuclear protein; ATP-binding; DNA-binding;  
KW Repeat.  
FT DOMAIN 191 199 POLY-GLU.  
FT DOMAIN 203 276 CHROMO 1.  
FT DOMAIN 304 365 CHROMO 2.  
FT NP\_BIND 415 422 ATP (POTENTIAL).  
FT SITE 524 527 DEAH BOX.  
FT DOMAIN 955 980 GLU-RICH.  
FT CONFLICT 68 68 I -> M (IN REF. 1).  
FT CONFLICT 133 133 S -> R (IN REF. 1).  
FT CONFLICT 225 225 D -> G (IN REF. 1).  
FT CONFLICT 339 339 A -> S (IN REF. 1).  
FT CONFLICT 385 385 I -> M (IN REF. 1).  
FT CONFLICT 479 489 NIREYEFYLSL -> ILES M (IN REF. 1).  
SQ SEQUENCE 1373 AA; 158547 MW; 7ABAF7BD4C2503B CRC64;  
  
Query Match 19.5%; Score 1402; DB 1; Length 1373;  
Best Local Similarity 28.0%; Pred. No. 1.5e-63;  
Matches 395; Conservative 273; Mismatches 467; Indels 274; Gaps 51;  
  
Qy 23 SDDDFVPKKDRTFEQVEALVRTDAKENACQAGESTNLVSCNTCTYAFHAKCLVPPLKD 82  
Db 145 SDDEDAYDKRRHF--SINSAGTEIRTSLSRSGKSVN-----YNEQEFYDDFED 192  
  
Qy 83 ASVENWRCPECVSPNLNEIDKILDCEMRPTKSSEQSSDAEPKPIFVKQYLVKWKGLSYLH 142  
Db 193 ---EEEEVEEQVE--EEVEPIDFVLNHRKRD--AQDDPKSSY--QYLIKQWEVSHLH 243  
  
Qy 143 CSWVPEKEFQ--KAYKSNHRLKTRVNNFHRQ-----MESFNSEDDFVAIR----- 186  
Db 244 NTWEDYSTLSSVRGYK-----KVDNYIKQNIYDREIREDPPTTFEDIEALDIERERK 296  
  
Qy 187 ----PEWTTVDRIACR-EEDGELEYLVKYSDECYWESESDI--STFQNEIQRF-K 238  
Db 297 NMLFEEYKIVERIVASETNEEGKTEYFVKWRQLPYDNCWE-DADVIYSMAPEVYQFLQ 355  
  
Qy 239 DVNSRTRRSKDVHKNRPDPQQDHTPEFLK-GLLHPYQLEGLNFLRFSWSKQTHVILA 297  
Db 356 RENSPLYLPYKGVFYNTRP-PYRKLEKQPSYIKGGEIRDFQLTGINWMAYLWHRNENGILA 414  
  
Qy 298 DEMGLGKTIQSIALLASLFE--ENLIPHLVIAPLSTLRNWEREFATWAPQMNVMYFGTA 355  
Db 415 DEMGLGKTVQVCFLSYLVHSLKQHGPFLLVPLSTVPAWQETLANWTDPDLSICYTGT 474  
  
Qy 356 QARAVIREHEFYLSKDQKKIKKKSGQISSESKQKRIKFDVLLTSYEMINLDSAVLKPIK 415  
Db 475 ESRANIREYEFYLSNRSK-----LKNFILLTTYEYILKDKQELANNIR 517  
  
Qy 416 WECMIVDEGHRLKNKDSKLFESSLTQYSSNHRILLTGTPLONNLDELFMLMHFLDAGKFGS 475  
Db 518 WQYLAIIDEAHLRLKNSESSLYETLSQFRTANRLITGTPLQNNLKELASLVNFLMPGKFYI 577  
  
Qy 476 LEEFQEEFKDINQEEQISRLHKMLAPHLRLRRVKDVMDMPKPKELILRLVDLSSLOKEYY 535





Db 174 KRWVPSLRVICFVGDKDARAAFIRDEMPG-----EWDVCVT 210  
Qy 400 SYEMINLDSAVLKPIKWECMIVDEGHLKNDKSLFSSLTQYSSNHRILLTGTPLQNNLD 459  
Db 211 SYEMVIKESVFKFHWRYLVIDEAHRKNEKSKLSEIVREFSTNRLLLTGTPLQNNLH 270  
Qy 460 ELFMLMHFLDAGKFGSLEEFQEEFKDIN---QEEQISRLHKMLAPHLRRVKDVMKDMP 516  
Db 271 ELWALLNELLDPVFNADDSDSWFTKNCLGDQKLVRLHAVLKPFLRRRIKTDVEKSLP 330  
Qy 517 PKKELILRVDLSSLQKEYYKAIFTRNYQVLTKKGGA-QISLNNIMMELRKVCCHPYMLEG 575  
Db 331 PKKEIKIYLGLSKMQREWYTKILMKDIDVLNSSGKMDKMLNLMQLRKCCNHPYLPDG 390  
Qy 576 VEPVIHDANEAFKQLLESCGKLQLDKMMVKLKEQHRVLIYTFQPHMLDLLEDYCTHKK 635  
Db 391 AEPGPPYTTD--EHIVSNSGKMVVLDKLLAKLKEQGSRVLFQWTRLLDILEDYCMWRG 448  
Qy 636 WQYERIDGKVGGAERQIR-----IDRFNAKNSNKECFLLSTRAGGLGINLATA 683  
Db 449 YEYCRLDGQTPHEEREDKFLVEFLQOREAIEAFNAPNSSKFIFMLSTRAGGLGINLASA 508  
Qy 684 DTVIIYDSDNPHADLOAMARAHRLQOTNKVMYIRLINRGTEERMMLTKKKMVLHLV 743  
Db 509 DVVILYDSDNPNQVDLOAMDRAHRIGQKKPVRVRLITDNTVEERIVERAEIKRLDSIV 568  
Qy 744 V--GKLKTQIN--QEELDDIIRYGSKELFASEDDGAGSKGIHYDDAAIDKLLDRDLV 798  
Db 569 IQQRLIDQRSNKLAKHEELQMIHGHATHVFASKSE-----LTDEDITTILERGEK 620  
Qy 799 EAEVVS-----VDDEENGFLKAFKVANFEYIDENEAAALEAQRVAAESK 843  
Db 621 KTAEMNERLQKMGESSLRFRMDIEQ-----SLYKFEGEDYREKQKLGME--WIEPPKR 673  
Qy 844 SSAGNSDRASYWEELLKDK-----FELHQAEEEL---NAL 874  
Db 674 ERKANYAVDAYFREALRVSEPKIPKAPRPPKQPNVQDFQFFPRLFELLEKEILYYRKT 733  
Qy 875 G-KRKRSRKQLVSIEDDLAGLEDVSS-DGDESYEAESTDGEAAGQGVQTRRRPYRRKGR 932  
Db 734 GYKVPRN-----PDIPNALAQREEQKKIDGAELTPETE----- 769  
Qy 933 DNLEPTPLMEGEGRSFRVLGFNQSRRAIFVQTLMRYGAGNFDWKEFVPRLKQKTFEINE 992  
Db 770 ---EKEKLLTQGFNTWTKRDFNQ-----FIKANEKYGRDDID--NIAREVEGKSPEEVME 819  
Qy 993 YGILFLKHIAEEIDENSPTFSDGVPRKGLRIEDVLVRIALLILVQEKVKFVEDHPGKPVF 1052  
Db 820 YSAVFWERCNELQD-----IEKIMAQIERGEARIQRRISIK---KALDAKIARYKA 867  
Qy 1053 PSRILERFGLRSGKIWKKEEHDKIMIRAVLKHGYGR 1088  
Db 868 PFHQLRIQYGTSGKKNYTEEDRFLICMLHMKMGFDR 903

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein.- protein search, using sw model

Run on: September 14, 2004, 01:52:16 ; Search time 162 Seconds  
(without alignments)  
2695.538 Million cell updates/sec

Title: US-10-049-137-2  
Perfect score: 7187  
Sequence: 1 MSSSLVERLIRSDRKPVYNL.....VDVKMEEAKEEKPKNMVVD 1384

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL 25:\*
- 1: sp\_archea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phage:\*
  - 10: sp\_plant:\*
  - 11: sp\_rodent:\*
  - 12: sp\_virus:\*
  - 13: sp\_vertebrate:\*
  - 14: sp\_unclassified:\*
  - 15: sp\_rvirus:\*
  - 16: sp\_bacteriap:\*
  - 17: sp\_archheap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7187	100.0	1384	10 Q9S775	Q9s775 arabidopsis
2	3114	43.3	1067	10 Q9SZ57	Q9sz57 arabidopsis
3	2105.5	29.3	1954	4 Q8TD10	Q8tdi0 homo sapien
4	2080	28.9	1937	4 Q8IXZ5	Q8ixz5 homo sapien
5	2076	28.9	1893	13 Q7ZWN3	Q7zwn3 xenopus lae
6	2066	28.7	2000	4 Q9Y4I0	Q9y4i0 homo sapien
7	2048.5	28.5	1829	5 Q19815	Q19815 caenorhabdi
8	1934.5	26.9	1388	4 Q9UFR9	Q9ufr9 homo sapien
9	1805	25.1	1045	11 Q8BM83	Q8bm83 mus musculu
10	1578	22.0	1808	13 Q42142	Q42142 gallus gall
11	1572	21.9	2259	10 Q7XIF9	Q7xif9 oryza sativ
12	1565.5	21.8	1738	10 Q9SI41	Q9si41 arabidopsis
13	1554	21.6	5322	5 Q9VPL9	Q9vpl9 drosophila
14	1553.5	21.6	1786	13 Q7ZT34	Q7zt34 poephila gu
15	1551	21.6	1806	13 Q800S6	Q800s6 poephila gu
16	1537	21.4	1461	5 O17909	O17909 caenorhabdi

17	1522.5	21.2	1883	5 Q24376	Q24376 drosophila
18	1505	20.9	2228	10 Q48579	Q48579 arabidopsis
19	1496.5	20.8	1081	11 Q8C7Q2	Q8c7q2 mus musculu
20	1490.5	20.7	981	13 Q9DGQ2	Q9dgg2 gallus gall
21	1456	20.3	918	13 Q9DGQ3	Q9dgg3 aegolius fu
22	1453	20.2	918	13 Q9DGQ1	Q9dgg1 nymphicus h
23	1437	20.0	918	13 Q9DGQ4	Q9dgg4 aegolius fu
24	1434	20.0	918	13 Q9DGQ0	Q9dgg0 nymphicus h
25	1345	18.7	1101	5 Q8SWV9	Q8swv9 drosophila
26	1332.5	18.5	2957	5 O61845	O61845 caenorhabdi
27	1290	17.9	1141	10 Q8LJJ7	Q8ljj7 oryza sativ
28	1263.5	17.6	1251	5 Q8SWP7	Q8swp7 encephalito
29	1258	17.5	3328	5 Q8IJG6	Q8ijg6 plasmodium
30	1224.5	17.0	1057	10 Q9SQU3	Q9squ3 arabidopsis
31	1220.5	17.0	1055	10 Q8RWY3	Q8rwy3 arabidopsis
32	1219.5	17.0	1137	10 Q9FRB8	Q9frb8 oryza sativ
33	1215	16.9	1122	10 Q93VU7	Q93vu7 oryza sativ
34	1195.5	16.6	891	11 Q8BIE8	Q8bie8 mus musculu
35	1173.5	16.3	1418	3 Q9P793	Q9p793 schizosacch
36	1157	16.1	730	4 Q9H9V7	Q9h9v7 homo sapien
37	1151.5	16.0	1052	4 O60264	O60264 homo sapien
38	1146.5	16.0	1051	11 Q91ZW3	Q91zw3 mus musculu
39	1143.5	15.9	1064	11 Q91Y58	Q91y58 mus musculu
40	1141.5	15.9	1051	11 Q925M9	Q925m9 mus musculu
41	1141	15.9	1102	5 Q7YYQ2	Q7yyq2 cryptospori
42	1139	15.8	1120	3 Q08773	Q08773 saccharomyc
43	1136	15.8	1046	11 Q8BSS1	Q8bss1 mus musculu
44	1136	15.8	1046	13 Q9DF71	Q9df71 xenopus lae
45	1132	15.8	1027	5 Q8SX14	Q8sx14 drosophila

ALIGNMENTS

RESULT 1				
Q9S775	ID	Q9S775	PRELIMINARY;	PRT; 1384 AA.
AC	Q9S775;			
DT	01-MAY-2000	(TREMBLrel. 13, Created)		
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)		
DT	01-OCT-2003	(TREMBLrel. 25, Last annotation update)		
DE	Chromatin remodeling factor CHD3.			
GN	PKL.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;			
OC	eurosids II; Brassicales; Brassicaceae; Arabidopsis.			
OX	NCBI_TaxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. Columbia;			
RX	MEDLINE=20040637; PubMed=10570159;			
RA	Ogas J., Kaufmann S., Henderson J., Somerville C.;			
RT	"PICKLE is a CHD3 chromatin-remodeling factor that regulates the			
RT	transition from embryonic to vegetative development in Arabidopsis."			
RL	Proc. Natl. Acad. Sci. U.S.A. 96:13839-13844(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. COLOMBIA;			
RA	Eshed Y., Baum S.F., Bowman J.L.;			
RT	"Distinct mechanisms promote polarity establishment in carpels of			
RT	Arabidopsis thaliana."			
RL	Cell 0:0-0(1999).			
DR	EMBL; AF185577; AAF13875.1; -.			
DR	EMBL; AF185578; AAF07084.1; -.			
DR	PIR; T52301; T52301.			
DR	GO; GO:0000785; C:chromatin; IEA.			
DR	GO; GO:0005634; C:nucleus; IEA.			
DR	GO; GO:0005524; F:ATP binding; IEA.			
DR	GO; GO:0008026; F:ATP dependent helicase activity; IEA.			
DR	GO; GO:0003682; F:chromatin binding; IEA.			
DR	GO; GO:0016787; F:hydrolase activity; IEA.			
DR	GO; GO:0006333; P:chromatin assembly/disassembly; IEA.			

DR	GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.	
DR	InterPro; IPR000953; Chromo.	
DR	InterPro; IPR001410; DEAD.	
DR	InterPro; IPR001650; Helicase_C.	
DR	InterPro; IPR000330; SNF2_N.	
DR	InterPro; IPR001965; Znf_PHD.	
DR	InterPro; IPR001841; Znf_ring.	
DR	Pfam; PF00385; chromo; 2.	
DR	Pfam; PF00271; helicase_C; 1.	
DR	Pfam; PF00628; PHD; 1.	
DR	Pfam; PF00176; SNF2_N; 1.	
DR	SMART; SM00298; CHROMO; 2.	
DR	SMART; SM00487; DEXDC; 1.	
DR	SMART; SM00490; HELICc; 1.	
DR	SMART; SM00249; PHD; 1.	
DR	SMART; SM00184; RING; 1.	
DR	PROSITE; PS50013; CHROMO 2; 2.	
DR	PROSITE; PS01359; ZF_PHD_1; 1.	
DR	PROSITE; PS50016; ZF_PHD 2; 1.	
KW	ATP-binding; Helicase; Hydrolase.	
SQ	SEQUENCE 1384 AA; 158404 MW; C4EDC75D7D973264 CRC64;	
Query Match		
Best Local Similarity 100.0%; Score 7187; DB 10; Length 1384;		
Matches 1384; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	1 MSSLVERLRIRSDRKPVYNLDDSDDDDFVPKKDRTFEQVEAIVRTDAKENACQAGESTN 60	
Db		
QY	61 LVSCNTCTYAFHAKCLVPPLKDasVENWRCPecvSPLNIDKILDCEMRPTKSSEQSSD 120	
Db		
QY	121 AEPKPIFVKQYLVKWKGLSYLHCSWVPEKFQKAYKSNHRLKTRVNNFHRQMESFNNSD 180	
Db		
QY	181 DFVAIRPEWTTVDRIACREEDGELEYLVKYKELSYDECYWESESDISTFQNEIQRFKDV 240	
Db		
QY	241 NSRTRRSKDVKHKNRPDRFQQFDHTPEFLKGLLHPYQLEGLNFLRFSWSKQTHVILADEM 300	
Db		
QY	301 GLGKTIQSIALLASLFEENLIPHLVIAPLSTLRNWEREFATWAPQNVVMYFGTAQARAV 360	
Db		
QY	361 IREHEFYLSKDQKIKKKKSGQISSSESKQRIKFDVLLTSYEMINLDSAVLKPIKWECEMI 420	
Db		
QY	421 VDEGHRLKNKDSKLFSSLTQYSSNHRILLTGTPLQNNLDELFLMHFLDAGKFGSLEEFQ 480	
Db		
QY	481 EEFKDINQEEQISRLHKWLAPHLLRRVKDVMKMPKPKELILRVDLSSLQKEYYKAIFT 540	
Db		
QY	541 RNYQVLTKKGGAQISLNNIMMELRKVCCHPYMLEGVEPVIHDANEAFKQLLESCKQLQL 600	
Db		
QY	601 DKMMVKLKEQHRVLIYTFQFHMLDLLEDYCTHKKQYERIDGKVGGAERQIRIDRFNAK 660	
Db		
QY	661 NSNKFCFLSTRAGGLGINLATADTVIIYDSDWNPHADLQAMARAHLGQTNKVMYIRLI 720	
Db		

QY	721 NRGTIERMMQLTKKKMWLEHLVVGKLKTONINQEELDDIIRYGSKELFASEDDEAGKSG 780	
Db		
QY	781 KIHYYDDAAIDKLLDRDLVEAEEVSVDDEENGFLKAFKVANFEYIDENEAALAEQORVAA 840	
Db		
QY	841 ESKSSAGNSDRASYWEELLDKDFELHQAEELNALGKRKRKQKLVSI EDDDLAGLEDVSS 900	
Db		
QY	901 DGDESYEAESTDGEAAGQGVQTRRPYRRKGRDNLEPTPLMEGEGRSFRLVGFNQSORAI 960	
Db		
QY	961 FVQITLMRYGAGNFDWKEFVPRLKQKTFFEEINEYGILFLKHIAEEIDENSPTFS DGPVKEG 1020	
Db		
QY	1021 LRIEDVLVRIALLILVQEKVKFVEDHPGKPVFPSPRILERPGLRSGKIWKKEHDKIMIRA 1080	
Db		
QY	1081 VLKHGYGRWQAI VDDKELGIGQELICKELNFPHISLSAAEQAGLQGGSGSNPGAQTNQ 1140	
Db		
QY	1141 NPGSVITGNNNASADGAQVNSMFYIRDQMORRLVEFVKRVLILLEKAMNYEYAEYYGLGG 1200	
Db		
QY	1201 SSSIPTEEPEAEPKIADTVGVSPIEVDDMLDGLPKTDPITSEEIMGAADVNNQARVEIA 1260	
Db		
QY	1261 QHYNQCKLLDENARESVAAYVNNQPPSTTKVNESFRALKSINGNINTILSITSDQSKSHE 1320	
Db		
QY	1321 DDTKPDLLNNVEMKDTAEETKPLRGGVVDLNVVEGEENIAEASGSVDVKMEEAKEEEKPKN 1380	
Db		
QY	1381 MVVD 1384	
Db		
RESULT 2		
Q9SZ57		
ID	Q9SZ57	PRELIMINARY; PRT; 1067 AA.
AC	Q9SZ57;	
DT	01-MAY-2000 (TrEMBLrel. 13, Created)	
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)	
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)	
DE	Hypothetical protein.	
GN	F11C18.100 OR AT4G31900.	
OS	Arabidopsis thaliana (Mouse-ear cress).	
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;	
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsi	
OX	NCBI_TaxID=3702;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Bevan M., Terry N., Ardiles W., Buysshaert C., Dasseville R.,	
RA	De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,	
RA	Villaroel R., Gielen J., Van Montagu M., Hoheisel J., Mewes H.W.,	
RA	Mayor K.F.X., Schueller C.;	
RL	Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.	
RN	[2]	
RP	SEQUENCE FROM N.A.	







DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)	
DE	Similar to chromodomain helicase DNA binding protein 4.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OX	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
NCBI_TaxID=9606;		
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Skin;	
RA	Strausberg R.;	
RL	Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; BC038596; AAH38596.1; -.	
DR	GO; GO:0000785; C:chromatin; IEA.	
DR	GO; GO:0005634; C:nucleus; IEA.	
DR	GO; GO:0005524; F:ATP binding; IEA.	
DR	GO; GO:0008026; F:ATP dependent helicase activity; IEA.	
DR	GO; GO:0003682; F:chromatin binding; IEA.	
DR	GO; GO:0003700; F:transcription factor activity; IEA.	
DR	GO; GO:0006333; P:chromatin assembly/disassembly; IEA.	
DR	GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.	
DR	InterPro; IPR000953; Chromo.	
DR	InterPro; IPR001410; DEAD.	
DR	InterPro; IPR002464; DEAH_box.	
DR	InterPro; IPR001650; Helicase_C.	
DR	InterPro; IPR000005; HTHArac.	
DR	InterPro; IPR000330; SNF2_N.	
DR	InterPro; IPR001965; Znf_PHD.	
DR	Pfam; PF00385; chromo; 1.	
DR	Pfam; PF00271; helicase_C; 1.	
DR	Pfam; PF00628; PHD; 2.	
DR	Pfam; PF00176; SNF2_N; 1.	
DR	SMART; SM00298; CHROMO; 2.	
DR	SMART; SM00487; DEXDc; 1.	
DR	SMART; SM00490; HELICc; 1.	
DR	SMART; SM00249; PHD; 2.	
DR	PROSITE; PS00598; CHROMO_1; 2.	
DR	PROSITE; PS00013; CHROMO_2; 2.	
DR	PROSITE; PS00690; DEAH_ATP_HELICASE; 1.	
DR	PROSITE; PS00041; HTH_ARAC_FAMILY_1; 1.	
DR	PROSITE; PS01359; ZF_PHD_1; 2.	
DR	PROSITE; PS00016; ZF_PHD_2; 2.	
KW	Helicase.	
SQ	SEQUENCE 1937 AA; 220474 MW; CDC16DF1E337E765 CRC64;	
Query Match 28.9%; Score 2080; DB 4; Length 1937;		
Best Local Similarity 34.9%; Pred.No.7.1e-108;		
Matches 520; Conservative 206; Mismatches 426; Indels 336; Gaps 44;		
QY	21 DSDDDDFVPKDRTFEQVEAIVRTDAKENACQACGESTNLVSCNTCTYAFHAKCLVPPL 80	
Db	423 DNSEGEIILBEVGGDLEED-----DHMEFCRVCKDGGELCCDTCPSYHIHCLNPPL 477	
QY	81 KDAVENWRCPCVSP--NEIDKILDCB--MRPTKSSEQSSDAE-----PKPIF---V 128	
Db	478 PEIPNGEWLRCRCTCPALKGVQKILIKWKQPPSPTPVPRPPDADPNTSPKPLEGRPE 537	
QY	129 KOYLVKWKGLSYLHCSWVPEKE-----FQKAYKSNHRLKTRVNNFHRQME----- 173	
Db	538 RQFFVKWQGSYWHCSWVSELOLELHCQVMFRNYQRKNDMDPEPSGDFGGDEKSRKRN 597	
QY	174 ---SFNNSEDDF--VAIRPEWTTVDRI--ACREEDGELEYLVKYKELSYDECYWESE--- 224	
Db	598 KDPKFAEMEERFYRYGIKPEWMMIHRILNHSVDKKGHVHYLIKWRDLPLYDAQSEWSEDE 657	
QY	225 -SDISTFQNEIQRFKDV-----NSRTRRSKDVVDHKNR-----DFQQFDHTPEFLK 270	
Db	658 IQDYDLFKQSYWNHRELMRGEGRPGKLLKVKLRKLERPPETPTVDPTVKYERQPEYLD 717	
QY	271 ---GLLHPYQLEGINFLRFSWSKQTHVILADEMGLKTIQSIALLASLFEE--NLIPHLY 325	
Db	718 ATGGTLHPYQMEGLNWLRFSWAQGTDTTILADEMGLKTVQTAFLYSLYKEGHSKGPFLV 777	
QY	326 IAPLSTLRNWEREFATWAPQMNVVMYFGTAQARAVIREHEFYLSKDQKIKKKKSGQISS 385	

Db	778 SAPLSTIINWEREFEMWAPDMYVVTVYVGDKDSRAIRENEFEFEDNAIRGGKKASRM--- 834	
QY	386 ESKQKRIKFDVLLTSYEMINLDSAVLKPIKWECMIVDEGHRLKNKDSKLFSSLTQYSSNH 445	
Db	835 -KKEASVKFHVLLTSYELITIDMAILGSDIWACLIIVDEAHLKNNQSKFFRVLNGYSLQH 893	
QY	446 RILLTGTPLQNNLDELFMHFLDAGKFGSLEEFQEEFKDINQEEQISRLHKLAPHLR 505	
Db	894 KLLLTGTPLQNNLEELFHLNFLTPTPERFHNLEGFLEEFADIAKEDQIKKLHMLGPHMLR 953	
QY	506 RVKQDVMDMPKPKELILRVDLSSLQKEYYKAIFTRNYQVLTTKGGA-QISLNNIMMELR 564	
Db	954 RLKADVFKMPSKTELIVRVELSPMQKYYKYILTRNFEALNARGGNQVSLNVVMDLK 1013	
QY	565 KVCCHPYMLEGVEPVIHDANEAFKQ-----LLESCGKLQLLDKMMVKLKEQGHVRL 615	
Db	1014 KCCNHPYLF---PVA--AMEAPKMPNGMYDGSALIRASGKLLLLQKMLKNLKEGHRVL 1067	
QY	616 IYTFQHMLDLLEDYCTHKKQYERIDGKVGAERQIRIDRFNAKNSNKFCFLSTRAGS 675	
Db	1068 IFSQMTKMLDLEDFLEHEGYKYERIDGGITGNMRQEAIDRFNAPGAQQFCFLSTRAGS 1127	
QY	676 LGINLATADTVIIYDSDNPHADLOAMARAHRLGQTNKMYIRLINRGTIERMMQLTKK 735	
Db	1128 LGINLATADTVIIYDSDNPHNDIQAFSRAHRIGQNKVMYIRFVTRASVEERITQVAKK 1187	
QY	736 KMVLEHLVVGK---LKTQININQEEELDDIIRYGSKELFASEDDEAGKSGK-----IHYD 785	
Db	1188 KMWLTHLVVRPGLSGTSGMSKQELDDILKFGTEELFKDEATDGGDNKEGEDSSVIHYD 1247	
QY	786 DAAIDKLLDRDLVEAEVSVDEEENGFLKAFKVANFEYIDENEAALAEQAVAAESKSS 845	
Db	1248 DKAIERLLDRNQDETEDEL--QGMNEYLSFSKVAQYVVRHEEMGEEEEVEEIIKQEEES 1305	
QY	846 AGNSDRASYWEELLKDFELHQAABELNALGKRKRSRKQL-----VSIEDDDLKAG----- 894	
Db	1306 VD----PDYWEKLLRRHHYEQQEDLARNLGKRIKQVNVNDGSDQEDRGVCGRPRPPPM 1361	
QY	895 -----LEDVSSDGDSEYAEASTDG--BAAGQGVQTGRRPYR---RKGR 932	
Db	1362 GRSTRAVGPAHLPSLPDWDQDSDNQSDYSVASEEGEDFDERSEAPRRPSRKGLRNDK 1421	
QY	933 DNLEPTPLMEGEGRSFRVLGPNQSQRAIFVQTLTRYG-----AGNFDWKEFVPRLKQKTF 987	
Db	1422 DKPLP-PLLARVGNIEVLGPNARQKAFNAIMRYGMPDQDAFTQW--LVDRLRGKSE 1478	
QY	988 EEINEYGILFLKHIAEEIDENSPTFSGVPKGLRIEDVLVRIALLILVQEKVKFVEDHP 1047	
Db	1479 KEFKAYVSLFMRHLCEPGADGAETFDAGVPREGLSRQHVLTRIGVMSLRKKVQEFE--- 1535	
QY	1048 GXPVFPRIERFPGLRSKGIWKEEHDKIMIRAVLKHGVRQWQAIVDDKELGIELICKE 1107	
Db	1536 -----HVNGRW----- 1541	
QY	1108 LNFPHISLSAAEQAGLQNGSGSNPGAQTNQNPQSVITGNNNNASADGAQVNSMFYIRD 1167	
Db	1542 -SMP--ELAEVEENKKMSQPGS--PSPKTPTPSTPGD-----TQNT----- 1578	
QY	1168 MQRRLVFEVKRVLLEKAMNYEYAEYYGLGSSSIPTPEEPAEPKIADTVGVSVFIEVD 1227	
Db	1579 -----PAPVPPAE----- 1586	
QY	1228 DEMLDGLPKTDPITSEEIMGAADVNNQARVEIAQHYNQMCKLLDENARESQAQVYVNNQPP 1287	
Db	1587 ----DGIK-----IEN----- 1594	
QY	1288 STKVNESFRALKSINGNI-NTILSITSDQSKSHEDDT-----KPDINNVMKDTAE- 1337	
Db	1595 SLKEESIEGEKEVKSTAPETAIECTQAPAPASEDEKVVVPEPPEGEKEKVAEVKERTEE 1654	
QY	1338 --ETKPLRGGVVDLNVVEGENIAEASGSVDVKMEEAKEEEKPKNMVV 1383	

Db

1655

PMETEP--KGAADVEKVE-EKSAIDLTPIVVEDKEEKKKEEKKVML

1699

RESULT 5

Q7ZWN3

Q7ZWN3

PRELIMINARY;

PRT; 1893 AA.

ID

Q7ZWN3;

AC

Q7ZWN3;

DT

01-JUN-2003 (TREMBLrel. 24, Created)

DT

01-JUN-2003 (TREMBLrel. 24, Last sequence update)

DT

01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE

Similar to chromodomain helicase DNA binding protein 4.

OS

Xenopus laevis (African clawed frog).

OC

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC

Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;

OC

Xenopodinae; Xenopus.

OX

NCBI\_TaxID=8355;

RN

[1]

RP

SEQUENCE FROM N.A.

RC

TISSUE=Embryo;

RA

Klein S., Strausberg R.;

RL

Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.

DR

EMBL; BC046866; AAH46866.1; -

DR

GO; GO:0000785; C:chromatin; IEA.

DR

GO; GO:0005634; C:nucleus; IEA.

DR

GO; GO:0005524; F:ATP binding; IEA.

DR

GO; GO:0008026; F:ATP dependent helicase activity; IEA.

DR

GO; GO:0003682; F:chromatin binding; IEA.

DR

GO; GO:0006333; P:chromatin assembly/disassembly; IEA.

DR

GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

DR

InterPro; IPR000953; Chromo.

DR

InterPro; IPR001410; DEAD.

DR

InterPro; IPR002464; DEAH\_box.

DR

InterPro; IPR001650; Helicase\_C.

DR

InterPro; IPR000330; SNF2\_N.

DR

InterPro; IPR001965; Znf\_PHD.

DR

Pfam; PF00271; helicase\_C; 1.

DR

Pfam; PF00628; PHD; 2.

DR

Pfam; PF00176; SNF2\_N; 1.

DR

SMART; SM00298; CHROMO; 2.

DR

SMART; SM00487; DEXDC; 1.

DR

SMART; SM00490; HELICC; 1.

DR

SMART; SM00249; PHD; 2.

DR

PROSITE; PS00013; CHROMO 2; 2.

DR

PROSITE; PS00690; DEAH\_ATP\_HELICASE; 1.

DR

PROSITE; PS50016; ZF\_PHD\_2; 2.

KW

Helicase.

SQ

SEQUENCE 1893 AA; 214687 MW; 5B03A501C48BC520 CRC64;

Query Match

28.9%;

Score 2076;

DB 13;

Length 1893;

Best Local Similarity

36.2%;

Pred. No. 1.2e-107;

Matches 526;

Conservative 206;

Mismatches 433;

Indels 288;

Gaps 47

QY

21

DDSDDDDFVPKKDRTFEQVEAIVRTDAKENACQAGGESTNLVSCNTCTYAFHAKCLVPPL

80

Db

423

DDLDDAVGDPEE-----DHMEFCRVCKDGGELCCDVCPSYHIHCLNPPL

470

QY

81

KDASVENWRCPECVSP--LNEIDKILDCEM----RPTKSSEQSSDAEPKPIF-----V

128

Db

471

PEIPNGEWLCPRCTCPPLKGKIQKILTWKGQPPPTPVSRPADSDPDAAAPLKPLEGRPE

530

QY

129

KQYLVKWKGLSYLHCSWVPEKE-----FQKAYKSN-----HRLKTRV

165

Db

531

REFFVKWHAMSYLHCSWVTELQLELHCQVMFRNYQRKNDMEPPAGDYGVDEEKSRRK

590

QY

166

NNFHRQMESFNNSDDF--VAIRPEWTTVDRI--ACREEDGELEYLVKYEKELSYDECYWE

222

Db

591

NKDPKYLE---MEEKFYRYGIKPEWMMTHRVNLNHSVDKKGYVHYLIKWRDLAYDQASWE

646

QY

223

S-ESDISTFQNETQRF-----KDVNSRTRRSKVDVHKR--NPRDFQ-----QFDHT

265

Db

647

AEESDIQDYDIYKQYWNHRELMCGDEGRPGKKVKVKLRKLRPPDTPAVDPTVKYDRQ

706

QY

266

PEFLK--GLLHPYQLEGLNFLRFSWSKQTHVILADEMGLGKTIQSIALLASLFEE--NL

320

Db	707	PDYLDATGGTLHPYQLEGLNWLRFWSWAQGTDTILADEMGLGKTQVTAVFLYSLYKEGHSK	766
QY	321	IPHLVIAPLSTLRNWEREFATWAPQMNVVMYFGTAQAAVIREHEFYLSKDQKKIKKKKS	380
Db	767	GPFLVSAPLSTIINWEREFEMWAPDMYVVTYVGDKDSRAVIRENEFSFEGNAIRGGKKAS	826
QY	381	GQISSESQKRIKFDVLLTSYEMINLDSAVLKPIKWECMIVDEGHRLKNKDSKLFSSLTQ	440
Db	827	RM-----KKEASVKFHVLLTSYELITIDTAVLGSIDWACLVDVAHRLKNNQSKFFRVLNG	882
QY	441	YSSNHRILLTGTPLQNNLDELFMLMHFLDAGKFGSLEEFQEEFKDINQEEQISRLLHKMLA	500
Db	883	YPLQHKLLLTGTPLQNNLEELFHLNFLTPERFNNLEGELEEFADIAKEDIKKLHDMLG	942
QY	501	PHLLRRVKDVMKMPPKKELILRVDSLSSLOKEYYKALFTRNYQVL-TKKGGAQISLNNI	559
Db	943	PHMLRRLKADVFKNMPSKTELTIVRVELSPMQKKYKFKILTFRNFEALNTRGGNQVSLNV	1002
QY	560	MMELRKVCCHPYMLEGVEPVTHDANEAFKQ-----LLESCGKLQLLDKMMVKLBQ	610
Db	1003	VMDLKCCCNHPYLF---PVA--AMEAPKMPNGMYDGSALIKGAGKFLLLQKMLRKLKDD	1056
QY	611	GHRVLIYTOFQHMLDLLEDYCTHKKWQYERIDGKVGGAERQIRIDRFNAKNSNKFCFLLS	670
Db	1057	GHRVIFSQMTKMLDLLEDMEHEGYKYERIDGGITGNMQEAIIDRFNAPGAQOFCFLS	1116
QY	671	TRAGGLGINLATADTVIIYDSDNWPHADLOAMARAHRLGQTNKVMYRLINRGTIERNM	730
Db	1117	TRAGGLGINLATADTVIYDSDNWPHNDIQAFSRAHRIQNRKKVMYRFVTRASVEERIT	1176
QY	731	QLTKKKMVLEHLVVGK--LKTQININQEEELDDIIRYGSKELFASEDDEAG--KSGK----	781
Db	1177	QVAKKKMMLTHLVVRPGLSGTGSMSKQELDDILKFGTEELFKDEATEGGENKEGEDISV	1236
QY	782	IHYDDAAIDKLLDRDLVEAEVSVDDDEENGFLKAFKVFANFEYIDENEAALAEAQRAAE	841
Db	1237	IHYDDKAIARLLDRNQDETEPEL-QLGMNEYLSSPKVAQYVVREEMGDEEVVREIIRK	1295
QY	842	SKSSAGNSDRASYWEELLKDKFELHQAEEELNALGKRKSRKQL---VSIEDDDLGLDED	897
Db	1296	QEESVD----PDYWEKLLRHYYEQQEDLARNLGKGRKIRKQVNYNDGSEQEDRD---WQD	1348
QY	898	VSSDGDSEYEAESTDG-EAACQGVQTRRRPYR---RKGRDNLEPTPLMEGEGRSFRVLGF	953
Db	1349	DQSDNQSDYSVASEEGDEDFDERSEAAARRPNRKGLRNDKDCKPLP-PLLARVGGNIEVLGF	1407
QY	954	NQQRALFVQTLMRYG-----AGNFDWKEFVPRLKQKTFFEEINEYGIILFKHIAEEDEN	1008
Db	1408	NARQKAFNALNAIMRYGMPPODAFTTQW--LVRDLRGSEKEBFKAYVSLFMRHLCPEGADG	1465
QY	1009	SPTFSDGVPKEGLRIEDVLRIALLILVOEKVKFVEDHPGKVPFPSPRILERFPGLRSGKI	1068
Db	1466	AETPADGVPREGLSRQHVLTRIGWMSLIRKKVQEFE-----	1501
QY	1069	WKEEHDKIMIRAVLKHGYGRWQAIVDKELGIQELICKELNFPPHISLSAAEQAGLQONG	1128
Db	1502	-----HVNGRW-----SMP--ELAEAEENKMASKID	1525
QY	1129	SGGSNPGAQTNQNPGSVITGNNNASADGAQVNSMFYRYDRMQRRLLVEFKRVLLLEKAMN	1188
Db	1526	S--PSPKTPTSTPGD--TQNTPTAPPASNEDST-----	1555
QY	1189	VEYABEYYGLGSSSIPTEEPFAEPKIADTVGVSFIEVDDEMLDGLPKTDPITSEEIMGA	1248
Db	1556	-----KSEVPVKEETDNEVKPT-----EENNETTQPTPDT-----QSP	1590
QY	1249	AVDNNQARVEIAQHYNQMCCKLLDENARESVOAYVNNQPPSTKVNESFRALKSINGINTI	1308
Db	1591	AAQTBEKKKEBPVQ-----QERTSPPPPPPPPPAPEPDMETDQ--KPEPEKQVDEV	1637
QY	1309	LSITSDQSKSHEDDTKPDLLNNVEM-----KDTAEE-TKPLR-----GGVV	1347



Db 1638 ITL-----DDKDDVQEVTLQNGETPKETAEBKTKASAAAQRFMFNIADGGFT 1688

QY 1348 DLNVV-EGEENIA 1359

Db 1689 ELHSLWQNEERAA 1701

RESULT 6

Q9Y4I0 PRELIMINARY; PRT; 2000 AA.

AC Q9Y4I0;

DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Zinc-finger helicas.

GN HZFH.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98351552; PubMed=9688266;

RA Aubry F.; Mattei M.G.; Galibert F.;

RT "Identification of a human 17p-located cDNA encoding a protein of the

RT Snf2-like helicas family.";

RL Eur. J. Biochem. 254:558-564(1998).

CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

CC -!- SIMILARITY: CONTAINS 1 'CHROMO' DOMAIN.

DR EMBL; U91543; AAC39923.1; -.

DR GO; GO:0005634; C:nucleus; NAS.

DR GO; GO:0008270; F:zinc ion binding; NAS.

DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.

DR InterPro; IPR000953; Chromo.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002464; DEAH\_box.

DR InterPro; IPR001650; Helicase\_C.

DR InterPro; IPR000330; SNF2\_N.

DR InterPro; IPR001965; Znf\_PHD.

DR InterPro; IPR001841; Znf\_ring.

DR Pfam; PF00385; chromo; 1.

DR Pfam; PF00271; helicase\_C; 1.

DR Pfam; PF00628; PHD; 2.

DR Pfam; PF00176; SNF2\_N; 1.

DR SMART; SM00298; CHROMO; 2.

DR SMART; SM00487; DEXDC; 1.

DR SMART; SM00490; HELICc; 1.

DR SMART; SM00184; RING; 2.

DR PROSITE; PS00598; CHROMO\_1; 1.

DR PROSITE; PS00690; DEAH\_ATP\_HELICASE; 1.

DR PROSITE; PS50016; ZF\_PHD\_2; 2.

KW ATP-binding; Helicase; Hydrolase; Nuclear protein.

SQ SEQUENCE 2000 AA; 226576 MW; B085A42A5AC7971F CRC64;

Query Match 28.7%; Score 2066; DB 4; Length 2000;

Best Local Similarity 33.5%; Pred. No. 4.6e-107;

Matches 520; Conservative 211; Mismatches 415; Indels 404; Gaps 41;

QY 52 CQACGESTNLVSCNTCTYAFHAKCLVPPPLKDasVENWRCPECVSP-NEIDKILDCEMR 109

Db 459 CRVCKDGGELCCDACISSYHIHCLNPPLDPDIPNGEWLCPRCTCPVLKGRVQKILHWRWG 518

QY 110 -----PTKSSEQSSDA-EPKPI---FVKQYLVKWKGLSYLHCSWVPEKEFQAYKSNH 159

Db 519 EPPVAVAPQADGNPDVPPPLQGRSREFFVKWVGLSYHCSWAKELQLEIFHLVWY 578

QY 160 RLKTRVNNFHRQME-SFNNSDD-----FVAIRPEWTTVDRI- 196

Db 579 RNYQRKNDMDPPPLDYGSGEDDGKSKRKYDPHYAEMEKEYRFGIKPEWMTVHRIIN 638

QY 197 ACREEDGELEYLVKYKELSYDECYW-ESESDISTFQNEIQ---RFKDVNSRTRRSKDV- 252

Db 639 HSDVKKGNHYHLVKWRDLPYDQSTWEDEMNIPEYEEHKQSYWRHRELIMGEDPAQPRKY 698

QY 253 KRNRDFQ-----QFDHTPEFLK---GLLHPYQLEGLNFLRFSWSKQTHVI 295

Db 699 KKKKELQGDGPPSPNTNDPTVKYETQPRFITATGGTLHMYQLEGLNWLRFSSWAQGTDTI 758

QY 296 LADENGLGKTIQSIALLASLEENLI--PHLVIAPLSTLRNWEREFATWAPQMNVMVYFG 353

Db 759 LADENGLGKTIQTIIVFLYSLYKEGHTKGPFLVSAPLSTIINWEREFQMWAPKFYVVVTYG 818

QY 354 TAOARAVIREHEFYLS---KDQKKIKKKSGQISSESKQRIKFDVLLTSEYEMINLDSA 409

Db 819 DKDSRAIIRENEFSFEDNAIKGKKAFAKMK-----REAQVKFHVLLTSYELITIDQA 870

QY 410 VLKPIKWECEMIVDEGHRLLKNKDSKLFSSLTQYSSNHRILLTGTPLONNLDELFLMHFLD 469

Db 871 ALGSIRWACLIVVDEAHRLLKNQSKFFRVNLGYKIDHKLLLTGTPLQNNLEELFHLNFLT 930

QY 470 AGKFGSLEEFQEEFKDINQEEQISRLHKMLAPHLLRRVVKDVMKMPKPKELILRVDLSS 529

Db 931 PERFNNLEGFLEEFADISKEDQIKKLHDLGPHMLRRLKADVFNKMPAKTELIVRVELSP 990

QY 530 LQKYYKAIFTRNQVLTKKGA-QISLNNIMMELRKVCCHPYMLEGVEPVIHDANEAFK 588

Db 991 MQKKYKYIILTRNFEALNSRGGNQVSLNIMMDLKKCNHPYLF---PVA--AMESPK 1044

QY 589 -----QLLESCGKLQLDDMMVKLKEQHRVLIYTFQHMLDLLEDYCTHKKWQYE 639

Db 1045 LPSGAYEGGALIKSSGKMLLQKMLRKLEQGHVLIPLFQMTKMLDLEDYCYKYE 1104

QY 640 RIDGKVGGAERQIRIDRFNAKNSNKFCFLSTRAGGLGINLATADTVIYDSWNPHADL 699

Db 1105 RIDGGITGALROEAIDRFNAPGAQQFCFLSTRAGGLGINLATADTVIIFSDWNPHNDI 1164

QY 700 QAMARAHRLGQTNKVMYRLINRGTIERMMQLTKKQNVLEHLVVGK---LKTQINQEE 756

Db 1165 QAFSRAHRIGQANKVMIYRFVTRASVEERITQVAKRKMLTHLVVRPGLGSKAGSMKQE 1224

QY 757 LDDIIRYGSKELFASDDDEAGK---SGKIHYDDAAIDKLLDRDLVEAEVSVDDDEENG 813

Db 1225 LDDILKFGTEELFKDENEGENKEEDSSVIHYDNEAIALLLDRNQDATEDTDV--QNMNEY 1282

QY 814 LKAFKVFANFEYIDENEAALAEARVAAESKSSAGNSDRASYWEELLKDKFELHQAELNA 873

Db 1283 LSSFVAQYVVRREDKIEIEREIIKQEE----NVD-PDYWEKLLRHHYEQQEDLARN 1336

QY 874 LGKRKRSRKQL--VSIEEDDLAGLEDVSSDGDESVEAESTDGEAAGQGVQTGRRPYRRK 930

Db 1337 LGKGRVVRKQVNYNDAQED-----QDNQSEYSVGSSEDEDFDERPEGRQSKRQLRNE 1391

QY 931 GRDNLEPTPLMEGEGRSFRVLGFNQSORAIFVQTLNRYG-----AGNFDWKEFVPRLKQK 985

Db 1392 -KDKPLP-PLLARVGGNIEVLGFNTRQRKAFNLNVRWGMPPQDAFTQW--LVRLRGK 1447

QY 986 TFEENEYGILFLKHIAEEIDENSPTFSDGVPKEGRIEDVLVRIALLILVQEKVKFVED 1045

Db 1448 TEKEFKAYVSLFMRHLCEPGADGSETFADGVPREGLSRQQVLTTRIGVMSLVKKKQVEFEH 1507

QY 1046 HPG-----SRILERF-----KPVFP----- 1053

Db 1508 INGRWSMPELMPDPSADSKRSSRASPSTKPTTPEASATNSPCTSKPATPAPSEKGEI 1567

QY 1054 -----SRILERF----- 1060

Db 1568 RTPLEKEEAENQEEKPEKNSRIGEKMETEADAPSPAPSLGERLEPRKIPLDEVPGVGE 1627

QY 1061 ----PGLR----- 1064

Db 1628 MEPEPGYRGDRKSATESTPGERGEKPLDQGEHRPEGETGDLGKREDYKGDRELAPG 1687

QY 1065 -----SGK---IWKEEHDKI 1076

Db 1688 PRDEPRSNGRREEKTENPRFMFNADGGFTLHTLWQNEERAATSSGKLNIEIWHRRHDYW 1747

QY 1077 MIRAVLKHGYGRWQAIVDDKELGIGIQLICKELNFPHISLSAAEQAGLQGQNGSGSNPGA 1136

Db 1748 LLAGIVLHGVARWQDIQNDAQFAI-----INEPF----- 1776

QY 1137 QTNQNPGSVITGNNNASADGAQVNSMFYYRDMQRRLVEFVKRVLLLEKAMNYE---YAE 1193

Db 1777 KTEANKGNFLEMKN-----KFLARRFKLLEQALVIEEQLRRA 1813

QY 1194 EYGLGGSSSIPTEEPEAEPKIADTVGVSVFIEVDDEMLDGLPKTDPITSEEIMGAAVDNN 1253

Db 1814 AYLNLSEQEPAHPAMALHARFAEAECLEASHQHSKESLAG-----NK 1855

QY 1254 QARVEIAQHYNQMKLLDENARE--SVQAYVNNQPP-STKVNESFRALKS 1300

Db 1856 PANAVLHKVLNQLELLSDMKADVTRLPATLSRIPPIAARLQMSERSILS 1905

RESULT 7

ID Q19815 PRELIMINARY; PRT; 1829 AA.

AC Q19815;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE F26F12.7 protein (LET-418).

GN F26F12.7 OR LET-418.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RN {1}

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RX MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Berks M.,

RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,

RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,

RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,

RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,

RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,

RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,

RA Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,

RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,

RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.,

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans.;"

RL Nature 368:32-38(1994).

RN {2}

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RA Wilson R., Bentley D., Gattung S.;

RT "The sequence of C. elegans cosmid F26F12.;"

RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.

RN {3}

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RA Waterston R.;

RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.

RN {4}

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RA Waterston R.;

RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.

RN {5}

RP SEQUENCE FROM N.A.

RA von Zelewsky T., Palladino F., Brunschwig K., Hajnal A., Mueller F.;

RT "The C. elegans Mi-2 chromatin-remodeling proteins function in vulval

RT cell fate determination.;"

RL Development 0:0-0(2000).

DR EMBL; U55373; AAC25894.1; --

DR EMBL; AF308445; AAG29838.1; --

DR PIR; T34239; T34239.

DR WormPep; F26F12.7; CE17716.

DR GO; GO:0000785; C:chromatin; IEA.

DR GO; GO:0005634; C:nucleus; IEA.

DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.

DR GO; GO:0003682; F:chromatin binding; IEA.

DR GO; GO:0016787; F:hydrolase activity; IEA.

DR GO; GO:0006333; P:chromatin assembly/disassembly; IEA.

DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

DR InterPro; IPR000953; Chromo.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002464; DEAH\_box.

DR InterPro; IPR001650; Helicase\_C.

DR InterPro; IPR000330; SNF2\_N.

DR InterPro; IPR001965; Znf\_PHD.

DR InterPro; IPR001841; Znf\_ring.

DR Pfam; PF00385; chromo; 1.

DR Pfam; PF00271; helicase\_C; 1.

DR Pfam; PF00628; PHD; 2.

DR Pfam; PF00176; SNF2\_N; 1.

DR SMART; SM00298; CHROMO; 2.

DR SMART; SM00487; DEXDC; 1.

DR SMART; SM00490; HELICc; 1.

DR SMART; SM00249; PHD; 2.

DR SMART; SM00184; RING; 2.

DR PROSITE; PS50013; CHROMO\_2; 2.

DR PROSITE; PS00690; DEAH\_ATP\_HELICASE; 1.

DR PROSITE; PS50016; ZF\_PHD\_2; 2.

KW ATP-binding; Helicase; Hydrolase.

SQ SEQUENCE 1829 AA; 209150 MW; 1A887E990C63B661 CRC64;

Query Match 28.5%; Score 2048.5; DB 5; Length 1829;

Best Local Similarity 32.9%; Pred. No. 3.9e-106;

Matches 518; Conservative 253; Mismatches 474; Indels 331; Gaps 44;

QY 40 EAIVRTDAKEN--ACQACGESTNLVSCNTCTYAFHAKCLVPLKDA-SVENWRCPEC--V 94

Db 306 EVVKEEPAKQNDKICKETENLLLCDSVCVSHAYCIDPPLTEVPKEETWSCPRCETV 365

QY 95 SPLNEIDKILDCEMR-----PTKSSEQGSSD-----AEPKPIFVKQYLVKWKGLSYL 141

Db 366 KPEHKIEKILCWRWKEIPYPEPLEAGKEASSDDAMLKPPRKMEPRREREFVKWKYLSYW 425

QY 142 HCSWVPEKBFQKAYKSNHRLKTRVNN-----FHRQMESFNNSDD-----FVAIR 186

Db 426 QCSWVSEMLLEVHFRMLILLYWRKNDSAPPEFEESVTSRHSNDPDYKLRERFYQYGIK 485

QY 187 PEWTTVDRIACRE-EDGELEYLVKYSDECYWE-SESDISTFQNEIQRF----- 237

Db 486 PEWQIHRINHQSYAKSQDYLKWKELSYDQATWERDDSNIANYEAIKYQWHRESK 545

QY 238 --KDVNSRTRR-----SKVDVHKRNPDRD---FQQFDHTPEFLK---GLLHPY 276

Db 546 LNEIDIPKNVQKMIAKHREAKGLPPKDEKKKKKKREKIDIRKKYEQPDYVTETGKCLHPY 605

QY 277 QLEGLNFIKFSWSKQTHVILADEMGLGKTIQSIALLASLFEENLI--PHLVIAPLSTLRN 334

Db 606 QLEGLNWLRLHCWSNGTDAILADEMGLGKTVQSLTFLYSLMKEGCKGPFLLAAPLSTIIN 665

QY 335 WEREFAFWAPQMNVMYFGTAQARAVIREHEFYLSKDQKKIKKKSGQISSSESQKRIKF 394

Db 666 WEREAEQWCPDFYVVTYVGLRDRVRLREHEFSFVEGAVRSGPKASKMKTEN----MKF 721

QY 395 DVLLTSYEMINLDSAVLKPIKWECKMIVDEGHRLLKNKSKLFSLLTQYSSNHRILLTGTP 454

Db 722 HVLLTSYETINMDKTISSIEWGALVVDEAHLKNNQSLFFKLNLENTIHYRVLLTGTP 781

QY 455 QNNLDELFLMHFLDAGKFGSLEEFQEEFKDINQEEQISRLHKKMLAPHLLRRVKKDVMDK 514

Db 782 QNNLEELFHLNFLSKERFNQLEAFTAEFNEISKEDQIEKHLNLLGPHMLRRLKADVL 841

QY 515 MPPKKELILRVDLSSLOKEYYKAIFTTNYQVL-TKKGGAQISLNNIMMELRKVCCHPYML 573













Db 1338 SDEDEPK-REYTAAGLALKEKYEKLARQKERI 1370

RESULT 12

Q9SI41

ID Q9SI41 PRELIMINARY; PRT; 1738 AA.

AC Q9SI41;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE SNF2 family (CHD1 subfamily) chromodomain protein.

GN AT2G13370.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RX MEDLINE=20083487; PubMed=10617197;

RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,

RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,

RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,

RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,

RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,

RA Copenhagen G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,

RA Salzberg S.L., Fraser C.M., Venter J.C.;

RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis

RT thaliana.";

RL Nature 402:761-768(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RA Lin X.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBDJ databases.

DR EMBL; AC007209; AAD28668.1; -.

DR PIR; C84507; C84507.

DR GO; GO:0000785; C:chromatin; IEA.

DR GO; GO:0005634; C:nucleus; IEA.

DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.

DR GO; GO:0003682; F:chromatin binding; IEA.

DR GO; GO:0016787; F:hydrolase activity; IEA.

DR GO; GO:0006333; P:chromatin assembly/disassembly; IEA.

DR InterPro; IPR000953; Chromo.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR001650; Helicase\_C.

DR InterPro; IPR000330; SNF2\_N.

DR Pfam; PF00385; Chromo; 1.

DR Pfam; PF00271; helicase C; 1.

DR Pfam; PF00176; SNF2\_N; 1.

DR SMART; SM00298; CHROMO; 1.

DR SMART; SM00487; DEXdc; 1.

DR SMART; SM00490; HELICC; 1.

DR PROSITE; PS50013; CHROMO 2; 2.

KW ATP-binding; Helicase; Hydrolase.

SQ SEQUENCE 1738 AA; 198696 MW; CE50069DD8C3B6CD CRC64;

Query Match 21.8%; Score 1565.5; DB 10; Length 1738;

Best Local Similarity 30.7%; Pred. No. 5.1e-79;

Matches 462; Conservative 237; Mismatches 485; Indels 321; Gaps 50;

QY 11 RSDRKPVNLDSDDDDFVPKKDRTFEQVEAIVR--TDAKENACQAGGESTNLV-----SC 64

Db 329 RQKRKTSYQDDSEED----SENDNDEGFRSLARRGTTLRQNN----GRSTNTIGQSSEV 380

QY 65 NTCTYAFHAKCLVPPLKDASVENWRCPECVSPLNE-----IDKILDCEMRPTKS 113

Db 381 RSSTRSVRKVSVESEDSIEDDGK-----NRKNQXDDIEEDADVIEKVLWHQLKGMGE 435

QY 114 SEQGSSDAEPKPIFVKQ-----YLVKWKGLSYLHCSWVPEFEFK----- 153

Db 436 DVQ-TNNKSTVPVLVSQLEFTEPDWNEMEFLIKWKQSHLHCQWKTLSDLQNLSGPKKVL 494

QY 154 --AYKSNHRLKTRVNNFHRQMESFNNSSEDDFVAIRPEWTTVDRIACR-EEDG-----ELE 206

Db 495 NYTKKVTEEIRYRTALSREEIEVNDVSKEMDLDIKQNSQVERIIADRIKSDGLGDVVPE 554

QY 207 YLVKYKELSYDECYWESESDISTFQNEIQRFKDVN-SRTRRSKDVDPHCR----- 254

Db 555 YLVKWQGLSYAEATWEKDVDFIAFAQVAIDEEKAREVSIQVQKQVEQRTTKGKGENSFSN 614

QY 255 -----NPRDFQQFDHTPEFL-KGLLHPYQLEGLNFLRFSWSKQTHVILADEMGLGKTI 306

Db 615 AELWLLFSVASLRKLDQPEWLIGGTLRDYQLEGLNFLVNSWLNDTNVILADEMGLGKTV 674

QY 307 QSIALLASLFEENLI--PHLVIAPLSTLRNWEREFATWAPQMVVYFGTAQARAVIREH 364

Db 675 QSVSMLGFLQNTQQIPGPFLVVVPLSTLANWAKEFRKWLPGMNIIVYVGTASREV---- 730

QY 365 EFYLSKDQKKIKKKSGQISSESQKRIKFDVLLTSYEMINLDSAVLKPIKWECEMIVDEG 424

Db 731 -----RNKTNDVHKVGRP--IKFNALLTYEVVVLKDKAVLSKIKWIYLMVDEA 776

QY 425 HRLKNKDSKLFSSLTQYSSNHRILLTGTPLONNLDLFMLMHFLDAGKFGSLEEFQSEFPK 484

Db 777 HRLKNSEAQLYTALLEFSTKNKLLITGTPLQNSVEELWALLHFLDPGKFKNKDEFVENYK 836

QY 485 DIN--QEEQISRLHKMLAPHLRLRRVKDKVMKMPKPKELILRVDSLSSQKEYYKAIFRN 542

Db 837 NLSSFNESELANLHLELRPHILRRVIKDVESKSLPPKIERILRVEMSPLOKQYKWLILRN 896

QY 543 YQVLTK-KGGAQISLNNIMMELRKVCCHPYMLE----GVEPVIHDANEAFKQLLESCGL 597

Db 897 FHDLNKGVGRGNQVSLNIIVVELKCCNHPFLFESADHGYGGDIND-NSKLDKIILSSGKL 955

QY 598 QLLDKMMVKLKEQGRVLIYTFQFHMLDLLEDYCTHKKWKQYERIDGKVGGAERQIRIDRF 657

Db 956 VILDKLLVRLRETKHRLVIFSQVMRMLDILAELYSLRGFQFRLDGSTKAELRQQAAMDF 1015

QY 658 NAKNSNKFCFLSTRAGGINLATADTVIYDSDWNPHADLQAMARAHRLGQTNKVMYI 717

Db 1016 NAPASDDFCFLSTRAGGINLATADTVVIFDSWNPQNDLQAMSAHRIGQQEVVNIY 1075

QY 718 PLINRGTIERMQLTKKKMVLHLVVGKL-----KTQNIHQEELDDIIRVGS 765

Db 1076 RFVTSKSVEEIILERAKRKMWLDHLVIQKLNAGERLKRETKKSGNFDKNELSAILRFGA 1135

QY 766 KELPASE-DDEAGSGKIHYYDDAAIDKLLDRDLVEAEVSVDDDEENGFLKAFKVANFEY 824

Db 1136 EELFKEDKNDEESKRLLSMD---IDEILER--AEQVEEKHTDETEHELLGAFKVANF-- 1188

QY 825 IDENEAALAEQRAVAESKSSAGNSDRASYWEELKDKFELHQAELNALGKR-KRSRKQ 883

Db 1189 -----CNAEDDGSFWSRWIKPD-SWVTAAE--ALAPRAARNTKS 1224

QY 884 LVSIEEDDLAGLEDVSSDGDESVEAESTDGEAAGQGVQTRRRPYRRKGRDNLEPTPLMBEG 943

Db 1225 YVDPSPDPRTSKR--KKXGSEPPE-----HTERSQKRRTKEYFVPSTPLLEG 1269

QY 944 -----EGRSFRVLGFNQSORAIFVQTLRMRYGAGNFDWKEFVPRLKQKTFFEEINEYGLFL 998

Db 1270 TSAQVRGWSYGNLPKRDAQR--FYRTVMKFGNHN-----Q-----QM 1303

QY 999 KHIAEEIDENSPFSDGVPKEGLRIEDVLVRIALLLVQEKVKF-----VEDHPGKP 1050

Db 1304 ACTAEEV-----GGVVEAAPEEAQVELFDALIDGCKESVETGNFEPKGPVLDFGVP 1355

QY 1051 VFPSRILERFPGLR-----SGKIWKEEHDKIMTRAVL 1082

Db 1356 VKANELLKRVOGLQLLSKRISRYNDPISQFRVLSYLKPSNWSKCGWNQIDARLLGLIL 1415

QY 1083 KHGYGRWQAI VDDKELGIGELICK-ELNFPHISLSAAEQAGLQGGSGGSPGQAQTNQN 1141

Db 1416 YHGFGNWEKIRLDESGLTKKIAPVELQHHTFL-----PRAPNLKE 1457



Best Local Similarity 29.8%; Pred. No. 1.2e-77;			Matches 454; Conservative 237; Mismatches 462; Indels 372; Gaps 60;		
QY	21	DDSDDDDFVPKKDRTE--EQVEAIVRTDAKENACQACGESTNLVSCNTCTYAFHAKCLVP	78		
Db	1742	DKMEVDVEGSKDKESKPEEQSETVKTEENSKAIEEDKSSTVL-----TADHAKEPETVLE	1797		
QY	79	PLK-----DASVENWRCPECVSPLNEIDKILDCEMRPTKSSEQ--	117		
Db	1798	KMEVDEKANDDQSAVSKAEGSDEKSTDDSNPEEATTEKNKESLIEGEKERVKEGESVK	1857		
QY	118	-----SSDAE--PKPIF--VKQYLVKWKGLSYLHCSWVPEKBFQKAYKSNHRLKTRVN	166		
Db	1858	KENDEKTEADMENKPEPVFIDVEEVFKYRNFSYLHCEWRTTEL--LKGDRRVAAKIR	1914		
QY	167	NFH----RQMESFNNSEDDFVAIRPEWTVDRILACR----BEDGEL--EYLVKYKELSY	216		
Db	1915	RFQOKSQQLNIFENIEDE--PFNQDFTVDRVLDMSVHTDETSGETTKHYLVKWKSLPY	1972		
QY	217	DECYWESESDISTFONEIQRFKDVNSRTRRSKDVHDKR--NPRDFQQFDHTPEFLKG--LLH	274		
Db	1973	EDCTWELEEDVD--NDKIEQYLRFNKIPQORSEWKSCKRPHPELWKKLEKTPVYKGGNSLR	2030		
QY	275	PYQLEGLNPLRFSWSKQTHVILADEMGLGKTIQSIALLASFEENLI--PHLVIAPLSTLR	333		
Db	2031	PYQLEGLNWLKFSWYNTHNCILADEMGLGKTIQSLTFVHSVVEYGIRGPFLLVIAPLSTIP	2090		
QY	334	NWEREFATWAPQMNVMYFGTAQARAVIREHEFYLSKQDKIKKKKSGQISSESKQKRIRK	393		
Db	2091	NWQREFEGWT--DMNVVVYHGSVTSQWIDQYEEYY-----KTESGKVLKEP----IK	2137		
QY	394	FDVLLTSYEMINLDSAVLKPIKWECMIVDEGHLRKNKDSKLFESSLTQYSSNHRILLTGP	453		
Db	2138	FNVLITTFEMIVTDYMDLKAFNWRLCVIDEAHRLKRNCKLLEGLRQLNLEHRVLLSGTP	2197		
QY	454	LQNNLDELMLMHFLDAGKFGSLEEFQEEFKDINQEEQISRLHKMLAPHLLRRVKKDVNK	513		
Db	2198	LQNNISELPSLLNLFLESPQSFSSQEEFMSFGLSRTEEEVNVKLQALLKPMMLRRLKDDVEK	2257		
QY	514	DMPPKKELLRVDLSSLQKEYYKAIFTRNYQVLTGKG--GAQI--SLNNIMMELRKVCCHP	570		
Db	2258	SLAPKEETIEVELTNIQKYYRGILBQNFSEFL--KKGTTSANIPNLMNTMELRKCCIH	2316		
QY	571	YMLEGVPEVI-----HDANEAFKQLLESCKGLQLLDKMMVKLKEQGHRLVIYTFQ	621		
Db	2317	YLLNGAEEQYQYDFKSHGEDPESYYKNLILSAGKMWLIDKLLPKIKANGHRVLIFSQMV	2376		
QY	622	HMLDLLEDYCTHKKQYERIDGKVGGAERQIRIDRFNAKNSNKFCFLSTRAGGLGINLA	681		
Db	2377	RCLDILEDLVYRKYPFERIDGRIRGNLRQEAIDRYSKPGSDRFVLLCTKAGGLGINLT	2436		
QY	682	TADTVIYDSDNPHADLOAMARAHRLGQTNKVMYIRLINRGTTIEERMQLTKKMWLEH	741		
Db	2437	AADTVIYDSDNPNQNDLQAQARCHRIGQKQKVKIYRLLCRNTYEREMFDKASMKLGLDK	2496		
QY	742	LVVGKLKTQ-----NINQEELEDDIIR---YGSKELFASEDDEAGSKGIHYDDAID	790		
Db	2497	AVLQSMNTQSGDKGNKQLSKEIEDLLKKGAYGA----VMDDDNAGDK----FCEEDID	2548		
QY	791	KLLDRDLVEABEVSVDDEENGFLKAFKVKANFEYIDENEAALAEQRAVAESKSSAGNSD	850		
Db	2549	SILKR---RTQVITMESEKGSTFKA-----SFAASGNRSDITID	2585		
QY	851	RASYWEELLK-----DKFELHQAEEELNALGKRKRSRKQLVLSIEEDDLAGLEDVSSGDGE	904		
Db	2586	DPDFWTKWAKKVDIDPDACERDETEDLVLSEPRRTQIKRY-----GHEDVMELNSE	2637		
QY	905	SYEAESTDGEACGQVQTRRPYRR-----KGRDNLEPTPLMEGEGRSFRVLGFNQSOR	958		
Db	2638	ESSNENS--EEGGIGLRSRRRKEKRDRAREKKGNDEYIPR-----ERDALAALGLEEIQY	2691		
QY	959	AIFVQT-----LMRYGAGNFDWKEFVPRLKQKTFEINEYGIPLKHIAREEIDENSP	1010		

Db	2692	GNWAKSECFKVEKGLLSFGWGR--WSELL-----ELGQF-----	2723		
QY	1011	TFSDGVPKEGLRIEDV--LVRIALLILVQ-----EKVK-FVEDHPGKVPFSPRILERFPG	1062		
Db	2724	-----KRGWRDVEDDCARIILLYCLQVYKGDEKIKTFIWD---LITPT-----	2764		
QY	1063	LRSWKIWKEEHDKIMIRAVLKHGYGRWQAIVDVDDKELGIGELICKELNFPHISLSAAEQAG	1122		
Db	2765	-EDGEVQKISR-----HSGLHNLVPRGRNG---GKSNKESTG	2798		
QY	1123	LQ--GQSGSGSNPGAQTNQNPQSGVITGNN-----NASADGAQVNS-----	1161		
Db	2799	VSTPAPGSASGSNSGNTTPAHKSSALDGSKDSSGINASAVAAVTSIHDPNHWSKKEKY	2858		
QY	1162	-----MFYYRDMQRRLV-EFVKK-----	1178		
Db	2859	DADAYLEGAYKKHLGRHANKVLLRVRMLY--IQHEVIGDFVQQIKONTPISELPIRPPP	2916		
QY	1179	-----RVLLEKAM---NYEYAEYY-----GLGGS-----SSI	1204		
Db	2917	TPDQVPSSWNPICCDKSLLVGTYKHGCEMYRQMRADPNLCFVTHVGAGAAGDDLAVTSL	2976		
QY	1205	PTEEPEAEPKIADTVGSFIEVDDEMLDGLPKTDPTITSEEIMGAADVNNQARVEIAQHYN	1264		
Db	2977	PINEDDANSKHDDG-----DEVDD---DGT'TTTKDSDSKLTGG--DNKDS-----	3017		
QY	1265	QMCKKLLD-----ENARESVOAYVNNQPPSTKVNESFRALKSINGNINTILS--ITSQSK	1317		
Db	3018	----LLDPERPSSGLDEESVAGSYPTVAAVEDATMWPSMQDLNTRLRRRVITAYQRN	3073		
QY	1318	SHEDDTKPDLLNVEMKDTABETKPL	1342		
Db	3074	YKKEELKLQ-QKAKLQALVSSSTPPL	3097		

RESULT 14  
Q7ZT34  
ID Q7ZT34 PRELIMINARY; PRT; 1786 AA.  
AC Q7ZT34;  
DT 01-JUN-2003 (TremBLrel. 24, Created)  
DT 01-JUN-2003 (TremBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)  
DE Chromo-helicase DNA-binding protein.  
GN CHDW.  
OS Poephila guttata (Zebra finch) (Taeniopygia guttata).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Passeriformes; Estrildidae;  
OC Estrildinae; Taeniopygia.  
OX NCBI\_TaxID=59729;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22584380; PubMed=12672961;  
RA Agate R.J., Grisham W., Wade J., Mann S., Wingfield J., Schanen C.,  
RA Palotie A., Arnold A.P.;  
RT "Neural, not gonadal, origin of brain sex differences in a  
RT gynandromorphic finch.";  
RL Proc. Natl. Acad. Sci. U.S.A. 100:4873-4878(2003).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Agate R.J., Arnold A.P.;  
RT "Sex Differences in Structure and Expression of the Sex Chromosome  
RT Genes CHDZ and CHDW in Zebra Finches.";  
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY217129; AA061781.1; -.  
DR EMBL; AY217130; AA061782.1; -.  
DR GO; GO:0000785; C:chromatin; IEA.  
DR GO; GO:0005634; C:nucleus; IEA.  
DR GO; GO:0000785; C:chromatin; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
DR GO; GO:0003682; F:chromatin binding; IEA.  
DR GO; GO:0008237; F:metallopeptidase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0006333; P:chromatin assembly/disassembly; IEA.







**This Page Blank (uspto)**